IN VIVO PRODUCTION OF CYCLIC PEPTIDES

This application claims the benefit of U.S.S.N. 60/187,130, filed March 6, 2000.

FIELD OF THE INVENTION

The present invention relates to methods and compositions for generating intracellular cyclic peptide and protein libraries.

BACKGROUND OF THE INVENTION

Combinatorial libraries of synthetic and natural products are important sources of molecular information for the development of pharmacologic agents. Linear peptide libraries, containing known and random peptide sequences, are particularly good sources of new and novel compounds for drug development because of the diversity of structures which can be generated. Drawbacks to linear peptide libraries are: (1) linear peptides are generally flexible molecules with entropic limitations on achieving productive biologically active conformations; (2) linear peptides are susceptibile to proteolytic enzymes; and, (3) linear peptides are inherently instable. For this reason, approaches utilizing conformational and topographical constraints to restrict the number of conformational states a peptide molecule may assume have been sought. See, for example, Hruby, (1982) Life Sci., 31:189; Hruby, et al., (1990) Biochem. J. 268:249.

Head-to-tail (backbone) peptide cyclization has been used to rigidify structure and improve *in vivo* stability of small bioactive peptides (see Camarero and Muir, (1999) J. Am. Chem. Soc., 121:5597-5598). An important consequence of peptide cyclization is retention of biological activity and/or the identification of new classes of pharmacological agents. Cyclic peptides have been reported that inhibit T-cell adhesion (Jois, et al. (1999) J. Pept. Res., 53:18-29), PDGF action (Brennand, et al. (1997) FEBS Lett., 413:70-74), and function as new classes of drugs (Kimura et al., (1997) J. Antibiot., 50:373-378; Eriksson, et al., (1989) Exp. Cell Res., 185:86-100).

Strategies for the preparation of circular polypeptides from linear precursors have been described. For example, a chemical cross-linking approach was used to prepare a backbone cyclized version of bovine pancreatic trypsin inhibitor (Goldenburg and Creighton (1983) J. Mol. Biol., 165:407-413). Other approaches include chemical (Camarero, et al., (1998) Angew. Chem. Int. Ed., 37:347-349; Tam and Lu (1998) Prot. Sci., 7:1583-1592; Camarero and Muir (1997) Chem. Commun., 1997:1369-1370; and Zhang and Tam (1997) J. Am. Chem. Soc. 119:2363-2370) and enzymatic (Jackson et al., (1995) J. Am. Chem. Soc., 117:819-820) intramolecular ligation methods which allow linear synthetic peptides to be efficiently cyclized under aqueous conditions. However, the requirement for synthetic peptide precursors has limited these chemical/enzymatic cyclization approaches to systems that are both ex vivo and limited to relatively small peptides.

One solution to this problem has been to generate circular recombinant peptides and proteins using a native chemical ligation approach. This approach utilizes inteins (*int*ernal pro*teins*) to catalyze head-to-tail peptide and protein ligation *in vivo* (see, for example, Evans, et al. (1999) J. Biol. Chem. 274:18359-18363; Iwai and Plückthun (1999) FEBS Lett. 459:166-172; Wood, et al. (1999) Nature Biotechnology 17:889-892; Camarero and Muir (1999) J. Am. Chem. Soc., 121:5597-5598; and Scott, et al. (1999) Proc. Natl. Acad. Sci. USA, 96:13638-13643).

Inteins are self-splicing proteins that occur as in-frame insertions in specific host proteins. In a self-splicing reaction, inteins excise themselves from a precursor protein, while the flanking regions, the exteins, become joined to restore host gene function. Inteins can also catalyze a trans-ligation self-splicing reaction. Approaches making use of the trans ligation reaction include splitting the intein into two parts and reassembling the two parts *in vitro*, each fused to a different extein (Southworth, et al., (1998) EMBO J. 17:918-926). A somewhat different approach uses an intein domain, and the reaction is then triggered with a thiolate nucleophile, such as DTT (Xu, et al., (1998) Protein Sci., 7:2256-2264).

The ability to construct intein fusions to proteins of interest has found several applications. For example, inteins can be used in conjunction with an affinity group to purify a desired protein (Wood, et al. (1999) Nature Biotechnology, 17:889-892). Circular recombinant fusion proteins have been created by cloning into a commercially available intein expression system (Camarero and Muir, (1999) J. Am. Chem. Soc., 121:5597-5598; Iwai and Plückthun (1999) FEBS Lett. 459:166-172; and Evans , et al. (1999) J. Biol. Chem. 274:18359-18363). In another approach, a mechanism for *in vivo* split intein-mediated circular ligation of peptides and proteins via permutation of the order of elements in the fusion protein precursor has been used to express cyclic products in bacteria (Scott, et al., (1999) Proc. Natl. Acad. Sci. USA, 96:13638-13643).

Cyclic peptide libraries have been generated in phage (Koivunen, et al., (1995) Biotechnology 13:265-70) and by using the backbone cyclic proteinomimetic approach (Friedler, et al., (1998) Biochemistry, 37:5616-22). Methods for modifying inteins for the purpose of creating cyclic peptides and/or proteins have been recently described (Benkovic, et al., WO 00/36093). It is an object of this invention to utilize intein function, derived from wild-type or mutant intein structures, to generate cyclic peptide libraries *in vivo*. The utilization of mutant intein structures for this purpose are of particular focus since these have been optimized for function in the specific context of an intein scaffold engineered to result in peptide/protein cyclization. Methods are described for generating, identifying, and utilizing mutants with altered splicing/cyclization activity for use with cyclic peptide/protein libraries. Intein-generated cyclic libraries are described for the identification of cyclic peptides/proteins capable of altering a given cellular phenotype. Accordingly, it is an object of the invention to provide compositions and methods useful in the generation of random fusion polypeptide libraries *in vivo*.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1A depicts head to tail protein cyclization by reconfigured/engineered intein.

Figure 1B depicts the mechanism of cyclization by reconfigured/engineered intein.

Figure 2A depicts intein catalyzed ligation by the Mxe GyrA intein. In it's normal configuration, intein catalyzed ligation joins the extein residues located at the junction points with each of the two intein motifs.

Figure 2B depicts the outcome of a motif reorganization resulting in the production of a cyclic peptide. Motif reorganziation involves providing intein B with its own translational start codon and placing intein B amino-terminal to intein A.

Figure 3A depicts the amino acid sequence of intein Ssp DnaB from *Synechocystis spp.*strain PCC6803.

Figure 3B depicts the amino acid sequence of intein Mxe GyrA from *Mycobacterium xenopi.*

Figure 3C depicts the amino acid sequence of intein Ceu ClpP from *Chlamydomonas eugametos.*

Figure 3D depicts the amino acid sequence of intein CIV RIR1 from Chilo iridescent virus.

Figure 3E depicts the amino acid sequence of intein Ctr VMA from *Candida tropicalis.*

Figure 3F depicts the amino acid sequence of intein Gth DnaB from *Guillardia theta*.

Figure 3G depicts the amino acid sequence of intein Ppu DnaB from *Porphyra purpurea*.

Figure 3H depicts the amino acid sequence of intein Sce VMA from *Saccharomyces cerevisiae*.

Figure 3I depicts the amino acid sequence of intein Mf1 RecA from *Mycobacterium flavescens*.

Figure 3J depicts the amino acid sequence of intein Ssp DnaE from *Synechocystis spp.*strain PCC6803.

Figure 3K depicts the amino acid sequence of intein Mle DnaB from *Mycobacterium leprae*.

Figure 3L depicts the amino acid sequence of intein Mja KlbA from *Methanococcus jannaschii*.

Figure 3M depicts the amino acid sequence of intein Pfu KlbA from *Pyrococcus furiosus*.

Figure 3N depicts the amino acid sequence of intein Mth RIR1 from *Methanobacterium thermoautotrophicum* (delta H strain).

Figure 3O depicts the amino acid sequence of intein Pfu RIR1-1 from *Pyrococcus furiosus*.

Figure 3P depicts the amino acid sequence of intein Psp-GBD Pol from *Pyrococcus spp.* GB-D.

Figure 3Q depicts the amino acid sequence of intein Thy Pol-2 from *Thermococcus hydrothermalis*.

Figure 3R depicts the amino acid sequence of intein Pfu IF2 from *Pyrococcus furiosus*.

Figure 3S depicts the amino acid sequence of intein Pho Lon from *Pyrococcus horikoshii* OT3.

Figure 3T depicts the amino acid sequence of intein Mja r-Gyr from *Methanococcus jannaschii*.

Figure 3U depicts the amino acid sequence of intein Pho RFC from *Pyrococcus horikoshii* OT3.

Figure 3V depicts the amino acid sequence of intein Pab RFC-2 from *Pyrococcus abyssi*.

Figure 3W depicts the amino acid sequence of intein Mja RtcB (Mja Hyp-2) from *Methanococcus jannaschii.*

Figure 3X depicts the amino acid sequence of intein Pho VMA from *Pyrococcus horikoshii* OT3.

Figure 4A depicts the amino acid sequence of a modified wild-type Ssp DnaB Intein. The DNA sequence is provided in Figure 4B.

Figures 5A and B depict the nucleotide and amino acid sequence of the intein Ssp DnaB J3 template used to generate intein mutants L7-J3, E6-J3, E9-J3, C11-J3 and B8-J3, with improved splicing efficiency. The J3 template carries a mutation which results in a amino acid change D to N at position 320. Thus, all mutants based on the J3 template are double mutants.

Figures 5C and D depict the nucleotide and amino acid sequence of intein mutant L7-J3. L7 has two mutations which result in amino acid changes: 1) D to N at position 320 and 2) R to K at position 389.

Figures 5E and F depict the nucleotide and amino acid sequence of intein mutant E6-J3. E6 has two mutations which result in amino acid changes: 1) D to N at position 320 and 2) I to V at position 34.

Figures 5G and H depict the nucleotide and amino acid sequence of intein mutant E9-J3. E9 has two mutations which result in amino acid changes: 1) D to N at position 320 and 2) T to A at position 36.

Figures 5I and J depict the nucleotide and amino acid sequence of intein mutant C11-J3. C11 has two mutations which result in amino acid changes: 1) D to N at position 320 and 2) S to P at position 23.

Figures 5K and L depict the nucleotide and amino acid sequence of intein mutant B8-J3. B8 has two mutations which result in amino acid changes: 1) D to N at position 320 and 2) K to R at position 369.

Figures 5M and N depict the nucleotide and amino acid sequence of intein mutant L7-wt, which was generated from an Ssp DnaB wild-type (wt) template. Mutants generated from the wt template carry a single mutation which effects splicing efficiency. L7-wt carries a single mutation which results in the amino acid change R to K at position 389.

Figures 5O and P depict the nucleotide and amino acid sequence of intein mutant C11-wt. C11-wt has a single mutation which result in the amino acid change S to P at position 23.

Figures 5Q and R depict the nucleotide and amino acid sequence of intein mutant E6-wt. E6-wt has a single mutation which result in the amino acid change I to V at position 34.

Figure 6 depicts the DNA sequence for a N-terminally fused GFP version of the Ssp DnaB intein.

Figure 7 depicts reporter proteins which can be used for the selection and/or detection of intein-based libraries.

Figure 8 depicts localization sequences which can be used to target cyclic peptide libraries.

Figure 9 depicts a random mutagenesis approach used in the optimization of intein cyclization function.

Figure 10 depicts a biotinylation approach for use in a yeast two hybrid system.

Figure 11 depicts a single chain antibody approach for use in a yeast two hybrid system.

Figure 12 depicts the fluorescent reporter system used to quantify intein cyclization. Figure 12 A depicts GFP split at the loop 3 junction and reversal of the translation order of the N- and C-terminal fragments. The termini are fused using a glycine-serine linker. The GFP is positioned within the Ssp DnaB intein cyclizationscaffold. Cyclized product reconstitutes both structure and fluorescence of GFP. In addition, splicing one-half of the myc epitope onto either side of the loop 3 junction allows for reconstruction of the myc epitope upon cyclization.

Figure 12B provides the amino acid sequence of DNAB intein cyclization scaffold with GFP.

Figure 12C depicts the mechanism of intein catalyzed cyclization of inverted loop 3 of GFP.

Figure 12D shows the results from a FACS analysis of the cyclization efficiency of wild-type Ssp DnaB intein in mammalian cells.

Figure 12E shows the results from a Western analysis of a Ssp DnaB catalyzed cyclization in mammalian cells.

Figure 12F shows the results form a native gel and the contribution to GFP fluorescence. The majority fo the fluorescence arises from the formation of cyclized GFP product, bands C and D.

Figure 13 illustrates a functional screen for isolating randomly-generated mutants with altered cyclization activity. Figure 13A depicts a functional screen for intein mutants with alteredd cyclization activity. Figure 13B depicts mutations modeled on the Mxe GyrA intein structure. Figure 13C depicts the sequence alignment of Mxe GyrA and Ssp DnaB inteins. Mutants are identified in shaded color. Figure 13D shows the results from a western analysis of isolated mutants. DnaB mutants E9-J3, E6-J3, C11-J3, L7-J3, and B8-J3 have cyclization efficiencies were are greater than the J3 starting intein template.

Figure 14 depicts intein-mediated excision/ligation in mammalian cells. Figure 14 A depicts constructs in which Ssp DnaB intein is inserted into loop 3 of GFP (i.e., GAB) or GFP with a C-terminal myc epitope. Figure 14B depicts constructs similar to those shown in 14A, except that the myc epitope half-sites are positioned onto the extreme ends of each splice junction. Figure 14C depicts Western blot analysis of lysates from transfected Phoenix cells. Lanes 3 and 4 demonstrate efficient splicing with only slight amounts of unspliced product detected.

Figures 15A-D depict a method for detecting cyclic peptides in mammalian cells. Figure 15A depicts an overview of the method in which cyclic peptides are detected in mammalian cells expressing a GFP fused intein scaffold with cyclic peptide inserts. Figures 15 B and C depict the MS analysis of mammalian cell lysates expressing the cyclic peptide products from RGD7 (15B) and RGD9 (15C). Figure 15D depicts an example of LC/MS fragmentation fingerprinting of the cyclic peptide product of an intein construct.

Figure 16 depicts the low energy conformers associated with cyclic peptide SRGDGWS.

Figure 17 depicts the low energy conformers associated with cyclic peptide SRGPGWS.

DETAILED DESCRIPTION OF THE INVENTION

Peptide libraries are an important source of new and novel drugs. However, a number of hurdles must be overcome in order to express and subsequently screen functional peptides and proteins in cells. Foremost amongst these hurdles is the need to retain biological activity of the peptides in a cellular environment. To overcome this problem, the present invention is directed to fusions of intein motifs and random peptides such that circular peptides are formed which retain biological activity.

Thus, generally, the present invention provides methods for generating libraries of cyclic peptides using inteins. Inteins are self-splicing proteins that occur as in-frame insertions in specific host proteins. In a self-splicing reaction, inteins excise themselves from a precursor protein, while the

flanking regions, the exteins, become joined via a new peptide bond to form a linear protein. By changing the N to C terminal orientation of the intein segments, the ends of the extein join, forming a cyclized extein. Figure 1 illustrates intein catalyzed joining of extein residues located at the junction points with each of the two intein motifs.

Because intein function is not strongly influenced by the nature of the extein polypeptide sequences located between them, standard recombinant methods can be used to insert random libraries into this position. Placement of these intein libraries into any number of delivery systems allows for the subsequent expression of unique cyclic peptides within individual cells. Such cells can then be screened to identify peptides of interest.

Accordingly, the present invention provides fusion polypetides comprising intein motifs and peptides.

By "fusion polypeptide" or "fusion peptide" or grammatical equivalents herein is meant a protein composed of a plurality of protein components, that while typically unjoined in their native state, are joined by their respective amino and carboxyl termini through a peptide linkage to form a single continuous polypeptide. "Protein" in this context includes proteins, polypeptides and peptides. Plurality in this context means at least two, and preferred embodiments generally utilize two components. It will be appreciated that the protein components can be joined directly or joined through a peptide linker/spacer as outlined below. In addition, as outlined below, additional components such as fusion partners including targeting sequences, etc. may be used.

The present invention provides fusion proteins of intein motifs and random peptides. By "inteins" , or "intein motifs", or "intein domains", or grammatical equivalents herein is meant a protein sequence which, during protein splicing, is excised from a protein precursor. Also included within in the definition of intein motifs are DNA sequences encoding inteins and mini-inteins.

Many inteins, are bifunctional proteins mediating both protein splicing and DNA cleavage. Such elements consist of a protein splicing domain interrupted by an endonuclease domain. Because endonuclease activity is not required for protein splicing, mini-inteins with accurate splicing activity can be generated by deletion of this central domain (Wood, et al., (1999) Nature Biotechnology, 17:889-892), hereby incorporated by reference.

Protein splicing involves four nucleophilic displacements by three conserved splice junction residues. These residues, located near the intein/extein junctions, include the initial cysteine, serine, or threonine of the intein, which intiates splicing with an acyl shift. The conserved cysteine, serine, or threonine of the extein, which ligates the exteins through nucleophilic attack, and the conserved C-

terminal histidine and asparagine of the intein, which releases the intein from the ligated exteins through succinimide formation. See Wood, et al., (1999) *supra*.

Inteins also catalyze a trans-ligation reaction. The ability of intein function to be reconstituted in *trans* by spatially separated intein domains suggests that reorganization of the self-splicing motifs can be used to produce peptides with a circular topology.

In a preferred embodiment, the translational order in which the N- and C-terminal intein motifs are normally synthesized within a polypetide chain is reversed. Generally, a reversal in the translational order in which the N- and C-terminal intein motifs are synthesized should not fundamentally change the enzymatic function of the intein. However, the location of the intervening peptide's amino and carboxy termini are altered in such a way that the product of the intein ligation reaction is no longer linear, but rather is cyclized. Figure 2 illustrates the outcome of a motif reorganization in which intein B has been given its own translational start codon and placed amino-terminal to intein A. To effectively express unique peptides in cells, fusion polypetides comprising a C-terminal motif, a peptide and a N-terminal motif are selected or designed for the production of random libraries of cyclic peptides *in vivo*.

In a preferred embodiment, the fusion polypeptide is designed with the primary sequence from the N-terminus comprising I_A-target-I_B. I_A is defined herein as the C-terminal intein motif, I_B is defined herein as the N-terminal intein motif and target is defined herein as a peptide. DNA sequences encoding the inteins may be obtained from a prokaryotic DNA sequence, such as a bacterial DNA sequence, or a eukaryotic DNA sequence, such as a yeast DNA sequence. The Intein Registry includes a list of all experimental and theoretical inteins discovered to date and submitted to the registry (http://www.neb.com/inteins/int_reg.html).

In a preferred embodiment, fusion polypeptides are designed using intein motifs selected from organisms belonging to the Eucarya and Eubacteria, with the intein Ssp DnaB (GenBank accession number Q55418) being particularly preferred. The GenBank accession numbers for other intein proteins and nucleic acids include, but are not limited to: Ceu ClpP (GenBank acession number P42379); CIV RIR1 (T03053); Ctr VMA (GenBank accession number A46080); Gth DnaB (GenBank accession number O78411); Ppu DnaB (GenBank accession number P51333); Sce VMA (GenBank accession number PXBYVA); Mf1 RecA (GenBank accession number not given); Mxe GyrA (GenBank accession number P72065); Ssp DnaE (GenBank accession number S76958 & S75328); and Mle DnaB (GenBank accession number CAA17948.1)

In other embodiments, inteins with alternative splicing mechanisms are preferred (see Southworth, et al., (2000) EMBO J., 19:5019-26). The GenBank accession numbers for inteins with alternative splicing mechanisms include, but are not limited to: Mja KlbA (GenBank accession number Q58191); and, Pfu KlbA (PF_949263 in UMBI).

In yet other embodiments, inteins from thermophilic organisms are used. Random mutagenesis or directed evolution (i.e. PCR shuffling, etc.) of inteins from these organisms could lead to the isolation of temperature sensitive mutants. Thus, inteins from thermophiles (i.e., Archaea) which find use in the invention are: Mth RIR1 (GenBank accession number G69186); Pfu RIR1-1 (AAB36947.1); Psp-GBD Pol (GenBank accession number AAA67132.1); Thy Pol-2 (GenBank accession number CAC18555.1); Pfu IF2 (PF_1088001 in UMBI); Pho Lon Baa29538.1); Mja r-Gyr (GenBank accession number G64488); Pho RFC (GenBank accession number F71231); Pab RFC-2 (GenBank accession number C75198); Mja RtcB (also referred to as Mja Hyp-2; GenBank accession number Q58095); and, Pho VMA (NT01PH1971 in Tigr).

Preferred fusion polypeptides of the invention increase the efficiency of the cyclization reaction by selecting or designing intein motifs with altered cyclization activity when expressed *in vivo*. In a preferred embodiment, the fusion polypeptides of the invention employ the DNA sequence encoding the *Synechocystis ssp.* strain PCC6803 DnaB intein. A particularly preferred fusion polypeptide structure is illustrated in Figure 4A and 4B.

In a preferred embodiment, fusion polypeptides are designed using mutant intein sequences with altered cyclization activity as described below. Preferred mutant intein sequences include, but are not limited, to those shown in Figure 5.

In a preferred embodiment, the fusion polypeptides of the invention comprise peptides. That is, the fusion polypeptides of the invention are translation products of nucleic acids. In this embodiment, nucleic acids are introduced into cells, and the cells express the nucleic acids to form peptides.] Generally, peptides ranging from about 4 amino acids in length to about 100 amino acids may be used, with peptides ranging from about 5 to about 50 being preferred, with from about 5 to about 30 being particularly preferred and from about 6 to about 20 being especially preferred.

In a preferred embodiment, the fusion polypeptides of the invention comprise random peptides. By "random peptides" herein is meant that each peptide consists of essentially random amino acids. Since generally these random peptides (or nucleic acids, discussed below) are chemically synthesized, they may incorporate any amino acid at any position. The synthetic process can be

designed to generate randomized proteins to allow the formation of all or most of the possible combinations over the length of the sequence, thus forming a library of randomized peptides.

In a preferred embodiment, the fusion polypeptides of the invention comprise peptides derived from a cDNA library.

The fusion polypeptide preferably includes additional components, including, but not limited to, reporter proteins and fusion partners.

In a preferred embodiment, the fusion polypeptides of the invention comprise a reporter protein. By "reporter protein" or grammatical equivalents herein is meant a protein that by its presence in or on a cell or when secreted in the media allow the cell to be distinguished from a cell that does not contain the reporter protein. As described herein, the cell usually comprises a reporter gene that encodes the reporter protein.

Reporter genes fall into several classes, as outlined above, including, but not limited to, detection genes, indirectly detectable genes, and survival genes. See Figure 6.

In a preferred embodiment, the reporter protein is a detectable protein. A "detectable protein" or "detection protein" (encoded by a detectable or detection gene) is a protein that can be used as a direct label; that is, the protein is detectable (and preferably, a cell comprising the detectable protein is detectable) without further manipulations or constructs. As outlined herein, preferred embodiments of screening utilize cell sorting (for example via FACS) to detect reporter (and thus peptide library) expression. Thus, in this embodiment, the protein product of the reporter gene itself can serve to distinguish cells that are expressing the detectable gene. In this embodiment, suitable detectable genes include those encoding autofluorescent proteins.

Detectable enzyme products resulting from the intein cyclization reaction may also be used to detect cells that are expressing the detectable product. Examples of enzymes which can be used include luciferase, β-galactosidase, β-lactamase, puromycin resistance protein, etc.

As is known in the art, there are a variety of autofluorescent proteins known; these generally are based on the green fluorescent protein (GFP) from *Aequorea* and variants thereof; including, but not limited to, GFP, (Chalfie, et al., "Green Fluorescent Protein as a Marker for Gene Expression," Science 263(5148):802-805 (1994)); enhanced GFP (EGFP; Clontech - Genbank Accession Number U55762)), blue fluorescent protein (BFP; Quantum Biotechnologies, Inc. 1801 de Maisonneuve Blvd. West, 8th Floor, Montreal (Quebec) Canada H3H 1J9; Stauber, R. H. Biotechniques 24(3):462-471

(1998); Heim, R. and Tsien, R. Y. Curr. Biol. 6:178-182 (1996)), enhanced yellow fluorescent protein (EYFP; Clontech Laboratories, Inc., 1020 East Meadow Circle, Palo Alto, CA 94303) and red fluorescent protein. In addition, there are recent reports of autofluorescent proteins from *Renilla* and *Ptilosarcus* species. See WO 92/15673; WO 95/07463; WO 98/14605; WO 98/26277; WO 99/49019; U.S. patent 5,292,658; U.S patent 5,418,155; U.S. patent 5,683,888; U.S. patent 5,741,668; U.S. patent 5,777,079; U.S. patent 5,804,387; U.S. patent 5,874,304; U.S patent 5,876,995; and U.S. patent 5,925,558; all of which are expressly incorporated herein by reference.

Preferred fluorescent molecules include but are not limited to green fluorescent protein (GFP; from *Aquorea* and *Renilla* species), blue fluorescent protein (BFP), yellow fluorescent protein (YFP),and red fluorescent protein (RFP).

In a preferred embodiment, the reporter protein is *Aequorea* green fluorescent protein or one of its variants; see Cody et al., Biochemistry 32:1212-1218 (1993); and Inouye and Tsuji, FEBS Lett. 341:277-280 (1994), both of which are expressly incorporated by reference herein. However, as is understood by those in the art, fluorescent proteins from other species may be used.

Accordingly, the present invention provides fusions of green fluorescent protein (GFP) and random peptides. By "green fluorescent protein" or "GFP" herein is meant a protein with at least 30% sequence identity to GFP and exhibits fluorescence at 490 to 600 nm. The wild-type GFP is 238 amino acids in length, contains a modified tripeptide fluorophore buried inside a relatively rigid β-can structure which protects the fluorophore from the solvent, and thus solvent quenching. See Prasher et al., Gene 111(2):229-233 (1992); Cody et al., Biochem. 32(5):1212-1218 (1993); Ormo et al, Science 273:1392-1395 (1996); and Yang et al., Nat. Biotech. 14:1246-1251 (1996), all of which are hereby incorporated by reference in their entirety). Included within the definition of GFP are derivatives of GFP, including amino acid substitutions, insertions and deletions. See for example WO 98/06737 and U.S. Patent No. 5,777,079, both of which are hereby incorporated by reference in their entirety. Accordingly, the GFP proteins utilized in the present invention may be shorter or longer than the wild type sequence. Thus, in a preferred embodiment, included within the definition of GFP proteins are portions or fragments of the wild type sequence. For example, GFP deletion mutants can be made. At the N-terminus, it is known that only the first amino acid of the protein may be deleted without loss of fluorescence. At the C-terminus, up to 7 residues can be deleted without loss of fluorescence; see Phillips et al., Current Opin. Structural Biol. 7:821 (1997)).

For fusions involving fluorescent proteins other than GFP, proteins with at least 24% sequence homology to YFP, RFP, BFP are included with the scope of the present invention.

In a preferred embodiment, intein A is fused to the N-terminus of GFP. The fusion can be direct, i.e. with no additional residues between the C-terminus of intein A and the N-terminus of GFP, or indirect; that is, intervening amino acids are inserted between the N-terminus of GFP and the C-terminus of intein A. See Figure 7.

In a preferred embodiment, intein B is fused to the C-terminus of GFP. As above for N-terminal fusions, the fusion can be direct or indirect.

In a preferred embodiment, the reporter protein is an indirectly detectable protein. As for the reporter proteins, cells that contain the indirectly detectable protein can be distinguished from those that do not; however, this is as a result of a secondary event. For example, a preferred embodiment utilizes "enzymatically detectable" reporters that comprise enzymes that will act on chromogenic, and particularly fluorogenic, substrates, to generate fluorescence, such as luciferase, β-galactosidase, and β-lactamase. Alternatively, the indirectly detectable protein may require a recombinant construct in a cell that may be activated by the reporter; for example, transcription factors or inducers that will bind to a promoter linked to an autofluorescent protein such that transcription of the autofluorescent protein occurs.

In a preferred embodiment, the indirectly detectable protein is a DNA-binding protein which can bind to a DNA binding site and activate transcription of an operably linked reporter gene. The reporter gene can be any of the detectable genes, such as green fluorescent protein, or any of the survival genes, outlined herein. The DNA binding site(s) to which the DNA binding protein is binding is (are) placed proximal to a basal promoter that contains sequences required for recognition by the basic transcription machinery (e.g., RNA polymerase II). The promoter controls expression of a reporter gene. Following introduction of this chimeric reporter construct into an appropriate cell, an increase of the reporter gene product provides an indication that the DNA binding protein bound to its DNA binding site and activated transcription. Preferably, in the absence of the DNA binding protein, no reporter gene product is made. Alternatively, a low basal level of reporter gene product may be tolerated in the case when a strong increase in reporter gene product is observed upon the addition of the DNA binding protein, or the DNA binding protein encoding gene. It is well known in the art to generate vectors comprising DNA binding site(s) for a DNA binding protein to be analyzed, promoter sequences and reporter genes.

In a preferred embodiment, the DNA-binding protein is a cell type specific DNA binding protein which can bind to a nucleic acid binding site within a promoter region to which endogenous proteins do not bind at all or bind very weakly. These cell type specific DNA-binding proteins comprise transcriptional activators, such as Oct-2 [Mueller et al., Nature 336(6199):544-51 (1988)] which e.g., is expressed in

lymphoid cells and not in fibroblast cells. Expression of this DNA binding protein in HeLa cells, which usually do not express this protein, is sufficient for a strong transcriptional activation of B-cell specific promoters, comprising a DNA binding site for Oct-2 (Mueller et al., supra).

In a preferred embodiment, the indirectly detectable protein is a DNA-binding/transcription activator fusion protein which can bind to a DNA binding site and activate transcription of an operably linked reporter gene. Briefly, transcription can be activated through the use of two functional domains of a transcription activator protein; a domain or sequence of amino acids that recognizes and binds to a nucleic acid sequence, i.e. a nucleic acid binding domain, and a domain or sequence of amino acids that will activate transcription when brought into proximity to the target sequence. Thus the transcriptional activation domain is thought to function by contacting other proteins required in transcription, essentially bringing in the machinery of transcription. It must be localized at the target gene by the nucleic acid binding domain, which putatively functions by positioning the transcriptional activation domain at the transcriptional complex of the target gene.

The DNA binding domain and the transcriptional activator domain can be either from the same transcriptional activator protein, or can be from different proteins (see McKnight et al., Proc. Natl. Acad. Sci. USA 89:7061 (1987); Ghosh et al., J. Mol. Biol. 234(3):610-619 (1993); and Curran et al., 55:395 (1988)). A variety of transcriptional activator proteins comprising an activation domain and a DNA binding domain are known in the art.

In a preferred embodiment the DNA-binding/transcription activator fusion protein is a tetracycline repressor protein (TetR)-VP16 fusion protein. This bipartite fusion protein consists of a DNA binding domain (TetR) and a transcription activation domain (VP16). TetR binds with high specificity to the tetracycline operator sequence, (tetO). The VP16 domain is capable of activating gene expression of a gene of interest, provided that it is recruited to a functional promoter. Employing a tetracycline repressor protein (TetR)-VP16 fusion protein, a suitable eukaryotic expression system which can be tightly controlled by the addition or omission of tetracycline or doxycycline has been described (Gossen and Bujard, Proc. Natl. Acad. Sci. U.S.A. 89:5547-5551; Gossen et al., Science 268:1766-1769 (1995)].

It is an object of the instant application to fuse intein amino acid sequences to DNA-binding/transcription activator proteins and/or to DNA-binding/transcription activator fusion proteins. N-terminal and C-terminal fusions are all contemplated. The site of fusion may be determined based on the structure of DNA-binding/transcription activator fusion protein, which are determined [e.g., TetR; see Orth et al., J. Mol. Biol. 285(2):455-61 (1999); Orth et al., J. Mol. Biol. 279(2):439-47 (1998);

Hinrichs et al., Science 264(5157):418-20 (1994); and Kisker et al., J. Mol. Biol. 247(2):260-80 (1995)].

In a preferred embodiment, the reporter protein is a survival protein. By "survival protein", "selection protein" or grammatical equivalents herein is meant a protein without which the cell cannot survive, such as drug resistance genes. As described herein, the cell usually does not naturally contain an active form of the survival protein which is used as a scaffold protein. As further described herein, the cell usually comprises a survival gene that encodes the survival protein.

The expression of a survival protein is usually not quantified in terms of protein activity, but rather recognized by conferring a characteristic phenotype onto a cell which comprises the respective survival gene or selection gene. Such survival genes may provide resistance to a selection agent (i.e., an antibiotic) to preferentially select only those cells which contain and express the respective survival gene. The variety of survival genes is quite broad and continues to grow (for review see Kriegler, Gene Transfer and Expression: A Laboratory Manual, W.H. Freeman and Company, New York, 1990). Typically, the DNA containing the resistance-conferring phenotype is transfected into a cell and subsequently the cell is treated with media containing the concentration of drug appropriate for the selective survival and expansion of the transfected and now drug-resistant cells.

Selection agents such as ampicillin, kanamycin and tetracycline have been widely used for selection procedures in prokaryotes [e.g., see Waxman and Strominger, Annu. Rev. Biochem. 52:825-69 (1983); Davies and Smith, Annu. Rev. Microbiol. 32:469-518 (1978); and Franklin, Biochem J., 105(1):371-8 (1967)]. Suitable selection agents for the selection of eukaryotic cells include, but are not limited to, blasticidin [Izumi et al., Exp. Cell Res., 197(2):229-33 (1991); Kimura et al., Biochim. Biophys. Acta 1219(3):653-9 (1994); Kimura et al., Mol. Gen. Genet. 242(2):121-9 (1994)], histidinol D [Hartman and Mulligan; Proc. Natl. Acad. Sci. U.S.A., 85(21):8047-51 (1988)], hygromycin [Gritz and Davies, Gene 25(2-3):179-88 (1983); Sorensen et al., Gene 112(2):257-60 (1992)], neomycin [Davies and Jimenez, Am. J. Trop. Med. Hyg., 29(5 Suppl):1089-92 (1980); Southern and Berg, J. Mol. Appl. Genet., 1(4):327-41 (19820], puromycin [de la Luna et al., Gene 62(1):121-6 (1988)] and bleomycin/phleomycin/zeocin antibiotics [Mulsant et al., Somat Cell. Mol. Genet. 14(3):243-52 (1988).

Survival genes encoding enzymes mediating such a drug-resistant phenotype and protocols for their use are known in the art (see Kriegler, supra). Suitable survival genes include, but are not limited to thymidine kinase [TK; Wigler et al., Cell 11:233 (1977)], adenine phosphoribosyltransferase [APRT; Lowry et al., Cell 22:817 (1980); Murray et al., Gene 31:233 (1984); Stambrook et al., Som. Cell. Mol. Genet. 4:359 (1982)], hypoxanthine-guanine phosphoribosyltransferase [HGPRT; Jolly et al., Proc. Natl. Acad. Sci. U.S.A. 80:477 (1983)], dihydrofolate reductase [DHFR; Subramani et al., Mol. Cell. Biol. 1:854 (1985); Kaufman and Sharp, J. Mol. Biol. 159:601 (1982); Simonsen and Levinson, Proc.

Natl. Acad. Sci. U.S.A. 80:2495 (1983)] aspartate transcarbamylase [Ruiz and Wahl, Mol. Cell. Biol. 6:3050 (1986)], ornithine decarboxylase [Chiang and McConlogue, Mol. Cell. Biol. 8:764 (1988)], aminoglycoside phosphotransferase [Southern and Berg, Mol. Appl. Gen. 1:327 (1982); Davies and Jiminez, supra], hygromycin-B-phosphotransferase [Gritz and Davies, supra; Sugden et al., Mol. Cell. Biol. 5:410 (1985); Palmer et al., Proc. Natl. Acad. Sci. U.S.A. 84:1055 (1987)], xanthine-guanine phosphoribosyltransferase [Mulligan and Berg, Proc. Natl. Acad. Sci. U.S.A. 78:2072 (1981)], tryptophan synthetase [Hartman and Mulligan, Proc. Natl. Acad. Sci. U.S.A. 85:8047 (1988)], histidinol dehydrogenase (Hartman and Mulligan, supra), multiple drug resistance biochemical marker [Kane et al., Mol. Cell. Biol. 8:3316 (1988); Choi et al., Cell 53:519 (1988)], blasticidin S deaminase [Izumi et al., Exp. Cell. Res. 197(2):229-33 (1991)], bleomycin hydrolase [Mulsant et al., supra], and puromycin-N-acetyl-transferase [Lacalle et al., Gene 79(2):375-80 (1989)],

In another preferred embodiment, the survival protein is blasticidin S deaminase, which is encoded by the bsr gene [Izumi et al., Exp. Cell. Res. 197(2):229-33 (1991)]. When transferred into almost any cell, this dominant selectable gene confers resistance to media comprising the antibiotic blasticidin S. Blasticidin S deaminase encoding genes have been cloned. They are used widely as a selectable marker on various vectors and the nucleotide sequences are available (e.g., see GenBank accession numbers D83710, U75992, and U75991).

It is an object of the instant application to fuse intein motif sequences to blasticidin S deaminase. N-terminal and C-terminal fusions are all contemplated. The site of fusion may be determined based on the structure of *Aspergillus terreus* blasticidin S deaminase, which has been determined [Nakasako et al., Acta Crystallogr. D. Biol. Crystallogr. 55(Pt2):547-8 (1999)]. Also, internal fusions can be done; see PCT US99/23715, hereby incorporated by reference.

In another preferred embodiment, the survival protein is puromycin-N-acetyl-transferase, which is encoded by the pac gene [Lacalle et al., Gene 79(2):375-80 (1989)]. When transferred into almost any cell, this dominant selectable gene confers resistance to media comprising puromycin. A puromycin-N-acetyltransferase encoding gene has been cloned. It is used widely as a selectable marker on various vectors and the nucleotide sequences are available (e.g., see GenBank accession numbers Z75185 and M25346).

It is an object of the instant application to fuse intein motif sequences puromycin-N-acetyl-transferase. N-terminal and C-terminal, dual N- and C-terminal and one or more internal fusions are all contemplated.

In a preferred embodiment, in addition to the intein motifs and peptides, the fusion polypeptides of the present invention preferably include additional components, including, but not limited to, fusion partners.

By "fusion partner" herein is meant a sequence that is associated with the fusion polypeptide that confers upon all members of the library in that class a common function or ability. Fusion partners can be heterologous (i.e. not native to the host cell), or synthetic (not native to any cell). Suitable fusion partners include, but are not limited to: a) targeting sequences, defined below, which allow the localization of the peptide into a subcellular or extracellular compartment; b) rescue sequences as defined below, which allow the purification or isolation of either the peptides or the nucleic acids encoding them; or c), any combination of a) and b).

In a preferred embodiment, the fusion partner is a targeting sequence. As will be appreciated by those in the art, the localization of proteins within a cell is a simple method for increasing effective concentration and determining function. For example, RAF1 when localized to the mitochondrial membrane can inhibit the anti-apoptotic effect of BCL-2. Similarly, membrane bound Sos induces Ras mediated signaling in T-lymphocytes. These mechanisms are thought to rely on the principle of limiting the search space for ligands, that is to say, the localization of a protein to the plasma membrane limits the search for its ligand to that limited dimensional space near the membrane as opposed to the three dimensional space of the cytoplasm. Alternatively, the concentration of a protein can also be simply increased by nature of the localization. Shuttling the proteins into the nucleus confines them to a smaller space thereby increasing concentration. Finally, the ligand or target may simply be localized to a specific compartment, and inhibitors must be localized appropriately.

Thus, suitable targeting sequences include, but are not limited to, binding sequences capable of causing binding of the expression product to a predetermined molecule or class of molecules while retaining bioactivity of the expression product, (for example by using enzyme inhibitor or substrate sequences to target a class of relevant enzymes); sequences signalling selective degradation, of itself or co-bound proteins; and signal sequences capable of constitutively localizing the peptides to a predetermined cellular locale, including a) subcellular locations such as the Golgi, endoplasmic reticulum, nucleus, nucleoli, nuclear membrane, mitochondria, chloroplast, secretory vesicles, lysosome, and cellular membrane; and b) extracellular locations via a secretory signal. Particularly preferred is localization to either subcellular locations or to the outside of the cell via secretion. See Figure 8.

In a preferred embodiment, the targeting sequence is a nuclear localization signal (NLS). NLSs are generally short, positively charged (basic) domains that serve to direct the entire protein in which they

occur to the cell's nucleus. Numerous NLS amino acid sequences have been reported including single basic NLS's such as that of the SV40 (monkey virus) large T Antigen (Pro Lys Lys Lys Arg Lys Val), Kalderon (1984), et al., Cell, 39:499-509; the human retinoic acid receptor-ß nuclear localization signal (ARRRRP); NFκB p50 (EEVQRKRQKL; Ghosh et al., Cell 62:1019 (1990); NFκB p65 (EEKRKRTYE; Nolan et al., Cell 64:961 (1991); and others (see for example Boulikas, J. Cell. Biochem. 55(1):32-58 (1994), hereby incorporated by reference) and double basic NLS's exemplified by that of the Xenopus (African clawed toad) protein, nucleoplasmin (Ala Val Lys Arg Pro Ala Ala Thr Lys Lys Ala Gly Gln Ala Lys Lys Lys Lys Leu Asp), Dingwall, et al., Cell, 30:449-458, 1982 and Dingwall, et al., J. Cell Biol., 107:641-849; 1988). Numerous localization studies have demonstrated that NLSs incorporated in synthetic peptides or grafted onto reporter proteins not normally targeted to the cell nucleus cause these peptides and reporter proteins to be concentrated in the nucleus. See, for example, Dingwall, and Laskey, Ann, Rev. Cell Biol., 2:367-390, 1986; Bonnerot, et al., Proc. Natl. Acad. Sci. USA, 84:6795-6799, 1987; Galileo, et al., Proc. Natl. Acad. Sci. USA, 87:458-462, 1990.

In a preferred embodiment, the targeting sequence is a membrane anchoring signal sequence. This is particularly useful since many parasites and pathogens bind to the membrane, in addition to the fact that many intracellular events originate at the plasma membrane. Thus, membrane-bound peptide libraries are useful for both the identification of important elements in these processes as well as for the discovery of effective inhibitors. The invention provides methods for presenting the randomized expression product extracelluarly or in the cytoplasmic space. For extracellular presentation, a membrane anchoring region is provided at the carboxyl terminus of the peptide presentation structure. The randomized expression product region is expressed on the cell surface and presented to the extracellular space, such that it can bind to other surface molecules (affecting their function) or molecules present in the extracellular medium. The binding of such molecules could confer function on the cells expressing a peptide that binds the molecule. The cytoplasmic region could be neutral or could contain a domain that, when the extracellular randomized expression product region is bound, confers a function on the cells (activation of a kinase, phosphatase, binding of other cellular components to effect function). Similarly, the randomized expression product-containing region could be contained within a cytoplasmic region, and the transmembrane region and extracellular region remain constant or have a defined function.

Membrane-anchoring sequences are well known in the art and are based on the genetic geometry of mammalian transmembrane molecules. Peptides are inserted into the membrane based on a signal sequence (designated herein as ssTM) and require a hydrophobic transmembrane domain (herein TM). The transmembrane proteins are inserted into the membrane such that the regions encoded 5' of the transmembrane domain are extracellular and the sequences 3' become intracellular. Of course, if these transmembrane domains are placed 5' of the variable region, they will serve to anchor it as an

intracellular domain, which may be desirable in some embodiments. ssTMs and TMs are known for a wide variety of membrane bound proteins, and these sequences may be used accordingly, either as pairs from a particular protein or with each component being taken from a different protein, or alternatively, the sequences may be synthetic, and derived entirely from consensus as artificial delivery domains.

As will be appreciated by those in the art, membrane-anchoring sequences, including both ssTM and TM, are known for a wide variety of proteins and any of these may be used. Particularly preferred membrane-anchoring sequences include, but are not limited to, those derived from CD8, ICAM-2, IL-8R, CD4 and LFA-1.

Useful sequences include sequences from: 1) class I integral membrane proteins such as IL-2 receptor β-chain (residues 1-26 are the signal sequence, 241-265 are the transmembrane residues; see Hatakeyama et al., Science 244:551 (1989) and von Heijne et al, Eur. J. Biochem. 174:671 (1988)) and insulin receptor β-chain (residues 1-27 are the signal, 957-959 are the transmembrane domain and 960-1382 are the cytoplasmic domain; see Hatakeyama, supra, and Ebina et al., Cell 40:747 (1985)); 2) class II integral membrane proteins such as neutral endopeptidase (residues 29-51 are the transmembrane domain, 2-28 are the cytoplasmic domain; see Malfroy et al., Biochem. Biophys. Res. Commun. 144:59 (1987)); 3) type III proteins such as human cytochrome P450 NF25 (Hatakeyama, supra); and 4) type IV proteins such as human P-glycoprotein (Hatakeyama, supra). Particularly preferred are CD8 and ICAM-2. For example, the signal sequences from CD8 and ICAM-2 lie at the extreme 5' end of the transcript. These consist of the amino acids 1-32 in the case of CD8 (MASPLTRFLSLNLLLLGESILGSGEAKPQAP; Nakauchi et al., PNAS USA 82:5126 (1985) and 1-21 in the case of ICAM-2 (MSSFGYRTLTVALFTLICCPG; Staunton et al., Nature (London) 339:61 (1989)). These leader sequences deliver the construct to the membrane while the hydrophobic transmembrane domains, placed 3' of the random peptide region, serve to anchor the construct in the membrane. These transmembrane domains are encompassed by amino acids 145-195 from CD8 (PQRPEDCRPRGSVKGTGLDFACDIYIWAPLAGICVALLLSLIITLICYHSR; Nakauchi, supra) and 224-256 from ICAM-2 (MVIIVTVVSVLLSLFVTSVLLCFIFGQHLRQQR; Staunton, supra).

Alternatively, membrane anchoring sequences include the GPI anchor, which results in a covalent bond between the molecule and the lipid bilayer via a glycosyl-phosphatidylinositol bond for example in DAF (PNKGSGTT**S**GTTRLLSGHTCFTLTGLLGTLVTMGLLT, with the bolded serine the site of the anchor; see Homans et al., Nature 333(6170):269-72 (1988), and Moran et al., J. Biol. Chem. 266:1250 (1991)). In order to do this, the GPI sequence from Thy-1 can be cassetted 3' of the variable region in place of a transmembrane sequence.

Similarly, myristylation sequences can serve as membrane anchoring sequences. It is known that the myristylation of c-src recruits it to the plasma membrane. This is a simple and effective method of membrane localization, given that the first 14 amino acids of the protein are solely responsible for this function: MGSSKSKPKDPSQR (see Cross et al., Mol. Cell. Biol. 4(9):1834 (1984); Spencer et al., Science 262:1019-1024 (1993), both of which are hereby incorporated by reference). This motif has already been shown to be effective in the localization of reporter genes and can be used to anchor the zeta chain of the TCR. This motif is placed 5' of the variable region in order to localize the construct to the plasma membrane. Other modifications such as palmitoylation can be used to anchor constructs in the plasma membrane; for example, palmitoylation sequences from the G protein-coupled receptor kinase GRK6 sequence (LLQRLFSRQD**CC**GN**C**SDSEEELPTRL, with the bold cysteines being palmitolyated; Stoffel et al., J. Biol. Chem 269:27791 (1994)); from rhodopsin (KQFRNCMLTSL**CC**GKNPLGD; Barnstable et al., J. Mol. Neurosci. 5(3):207 (1994)); and the p21 H-ras 1 protein (LNPPDESGPGCMS**C**K**C**VLS; Capon et al., Nature 302:33 (1983)).

In a preferred embodiment, the targeting sequence is a lysozomal targeting sequence, including, for example, a lysosomal degradation sequence such as Lamp-2 (KFERQ; Dice, Ann. N.Y. Acad. Sci. 674:58 (1992); or lysosomal membrane sequences from Lamp-1 (*MLIPIAGFFALAGLVLIVLIAYLI*<u>GRKRSHAGYQTI</u>, Uthayakumar et al., Cell. Mol. Biol. Res. 41:405 (1995)) or Lamp-2 (*LVPIAVGAALAGVLILVLLAYFIG*<u>LKHHHAGYEQF</u>, Konecki et la., Biochem. Biophys. Res. Comm. 205:1-5 (1994), both of which show the transmembrane domains in italics and the cytoplasmic targeting signal underlined).

Alternatively, the targeting sequence may be a mitochondrial localization sequence, including mitochondrial matrix sequences (e.g. yeast alcohol dehydrogenase III; MLRTSSLFTRRVQPSLFSRNILRLQST; Schatz, Eur. J. Biochem. 165:1-6 (1987)); mitochondrial inner membrane sequences (yeast cytochrome c oxidase subunit IV; MLSLRQSIRFFKPATRTLCSSRYLL; Schatz, supra); mitochondrial intermembrane space sequences (yeast cytochrome c1; MFSMLSKRWAQRTLSKSFYSTATGAASKSGKLTQKLVTAGVAAAGITASTLLYADSLTAEAMTA; Schatz, supra) or mitochondrial outer membrane sequences (yeast 70 kD outer membrane protein; MKSFITRNKTAILATVAATGTAIGAYYYYNQLQQQQRGKK; Schatz, supra).

The target sequences may also be endoplasmic reticulum sequences, including the sequences from calreticulin (KDEL; Pelham, Royal Society London Transactions B; 1-10 (1992)) or adenovirus E3/19K protein (LYLSRRSFIDEKKMP; Jackson et al., EMBO J. 9:3153 (1990).

Furthermore, targeting sequences also include peroxisome sequences (for example, the peroxisome matrix sequence from Luciferase; SKL; Keller et al., PNAS USA 4:3264 (1987)); farnesylation

sequences (for example, P21 H-ras 1; LNPPDESGPGCMSCK**C**VLS, with the bold cysteine farnesylated; Capon, supra); geranylgeranylation sequences (for example, protein rab-5A; LTEPTQPTRNQ**CC**SN, with the bold cysteines geranylgeranylated; Farnsworth, PNAS USA 91:11963 (1994)); or destruction sequences (cyclin B1; RTALGDIGN; Klotzbucher et al., EMBO J. 1:3053 (1996)).

In a preferred embodiment, the targeting sequence is a secretory signal sequence capable of effecting the secretion of the peptide. There are a large number of known secretory signal sequences which are placed 5' to the variable peptide region, and are cleaved from the peptide region to effect secretion into the extracellular space. Secretory signal sequences and their transferability to unrelated proteins are well known, e.g., Silhavy, et al. (1985) Microbiol. Rev. 49, 398-418. This is particularly useful to generate a peptide capable of binding to the surface of, or affecting the physiology of, a target cell that is other than the host cell, e.g., the cell infected with the retrovirus. In a preferred approach, a fusion polypeptide is configured to contain, in series, a secretion signal peptide-intein B motif-randomized library sequence-intein A. See Figure 8. In this manner, target cells grown in the vicinity of cells caused to express the library of peptides, are bathed in secreted peptide. Target cells exhibiting a physiological change in response to the presence of a peptide, e.g., by the peptide binding to a surface receptor or by being internalized and binding to intracellular targets, and the secreting cells are localized by any of a variety of selection schemes and the peptide causing the effect determined. Exemplary effects include variously that of a designer cytokine (i.e., a stem cell factor capable of causing hematopoietic stem cells to divide and maintain their totipotential), a factor causing cancer cells to undergo spontaneous apoptosis, a factor that binds to the cell surface of target cells and labels them specifically, etc.

Suitable secretory sequences are known, including signals from IL-2 (MYRMQLLSCIALSLALVTNS; Villinger et al., J. Immunol. 155:3946 (1995)), growth hormone (MATGSRTSLLLAFGLLCLPWLQEGSAFPT; Roskam et al., Nucleic Acids Res. 7:30 (1979)); preproinsulin (MALWMRLLPLLALLALWGPDPAAAFVN; Bell et al., Nature 284:26 (1980)); and influenza HA protein (MKAKLLVLLYAFVAGDQI; Sekiwawa et al., PNAS 80:3563)), with cleavage between the non-underlined-underlined junction. A particularly preferred secretory signal sequence is the signal leader sequence from the secreted cytokine IL-4, which comprises the first 24 amino acids of IL-4 as follows: MGLTSQLLPPLFFLLACAGNFVHG.

In a preferred embodiment, the fusion partner is a rescue sequence. A rescue sequence is a sequence which may be used to purify or isolate either the peptide or the nucleic acid encoding it. Thus, for example, peptide rescue sequences include purification sequences such as the His_6 tag for use with Ni affinity columns and epitope tags for detection, immunoprecipitation or FACS

(fluorescence-activated cell sorting). Suitable epitope tags include myc (for use with the commercially available 9E10 antibody), the BSP biotinylation target sequence of the bacterial enzyme BirA, flu tags, lacZ, GST, and Strep tag I and II.

Alternatively, the rescue sequence may be a unique oligonucleotide sequence which serves as a probe target site to allow the quick and easy isolation of the retroviral construct, via PCR, related techniques, or hybridization.

While the discussion has been directed to the fusion of fusion partners to the intein portion of the fusion polypeptide, the fusion partners may be placed anywhere (i.e. N-terminal, C-terminal, internal) in the structure as the biology and activity permits. In addition, it is also possible to fuse one or more of these fusion partners to the intein portions of the fusion polypeptide. Thus, for example, a targeting sequence (either N-terminally, C-terminally, or internally, as described below) may be fused to intein A, and a rescue sequence fused to the same place or a different place on the molecule. Thus, any combination of fusion partners and peptides may be made.

In a preferred embodiment, the invention provides libraries of fusion polypeptides. By "library" herein is meant a sufficiently structurally diverse population of randomized expression products to effect a probabilistically sufficient range of cellular responses to provide one or more cells exhibiting a desired response. Accordingly, an interaction library must be large enough so that at least one of its members will have a structure that gives it affinity for some molecule, protein, or other factor whose activity is of interest. Although it is difficult to gauge the required absolute size of an interaction library, nature provides a hint with the immune response: a diversity of 10^7-10^8 different antibodies provides at least one combination with sufficient affinity to interact with most potential antigens faced by an organism. Published in vitro selection techniques have also shown that a library size of 10^7 to 10^8 is sufficient to find structures with affinity for the target. A library of all combinations of a peptide 7 to 20 amino acids in length, such as proposed here for expression in retroviruses, has the potential to code for 20^7 (10^9) to 20^{20}. Thus, with libraries of 10^7 to 10^8 per ml of retroviral particles the present methods allow a "working" subset of a theoretically complete interaction library for 7 amino acids, and a subset of shapes for the 20^{20} library. Thus, in a preferred embodiment, at least 10^6, preferably at least 10^7, more preferably at least 10^8 and most preferably at least 10^9 different expression products are simultaneously analyzed in the subject methods. Preferred methods maximize library size and diversity.

In a preferred embodiment, libraries of all combinations of a peptide 3 to 30 amino acids in length are synthesized and analyzed as outlined herein. Libraries of smaller cyclic peptides, i.e., 3 to 4 amino acid in length, are advantageous because they are more constrained and thus there is a better chance

that these libraries possess desirable pharmocokinetics properties as a consequence of their smaller size. Accordingly, the libraries of the present invention may be one of any of the following lengths: 3 amino acids, 4 amino acids, 5 amino acids, 6 amino acids, 7 amino acids, 8 amino acids, 9 amino acids, 10 amino acids, 11 amino acids, 12 amino acids, 13 amino acids, 14 amino acids, 15 amino acids, 16 amino acids, 17 amino acids, 18 amino acids, 19 amino acids, 20 amino acids, 21 amino acids, 22 amino acids, 23 amino acids, 24 amino acids, 25 amino acids, 26 amino acids, 27 amino acids, 28 amino acids, 29 amino acids and 30 amino acids in length.

The invention further provides fusion nucleic acids encoding the fusion polypeptides of the invention. As will be appreciated by those in the art, due to the degeneracy of the genetic code, an extremely large number of nucleic acids may be made, all of which encode the fusion proteins of the present invention. Thus, having identified a particular amino acid sequence, those skilled in the art could make any number of different nucleic acids, by simply modifying the sequence of one or more codons in a way which does not change the amino acid sequence of the fusion protein.

Using the nucleic acids of the present invention which encode a fusion protein, a variety of expression vectors are made. The expression vectors may be either self-replicating extrachromosomal vectors or vectors which integrate into a host genome. Generally, these expression vectors include transcriptional and translational regulatory nucleic acid operably linked to the nucleic acid encoding the fusion protein. The term "control sequences" refers to DNA sequences necessary for the expression of an operably linked coding sequence in a particular host organism. The control sequences that are suitable for prokaryotes, for example, include a promoter, optionally an operator sequence, and a ribosome binding site. Eukaryotic cells are known to utilize promoters, polyadenylation signals, and enhancers.

The fusion nucleic acids are introduced into cells to screen for cyclic peptides capable of altering the phenotype of a cell. By "introduced into " or grammatical equivalents herein is meant that the nucleic acids enter the cells in a manner suitable for subsequent expression of the nucleic acid. The method of introduction is largely dictated by the targeted cell type, discussed below. Exemplary methods include $CaPO_4$ precipitation, liposome fusion, lipofectin®, electroporation, viral infection, etc. The fusion nucleic acids may stably integrate into the genome of the host cell (for example, with retroviral introduction, outlined below), or may exist either transiently or stably in the cytoplasm (i.e. through the use of traditional plasmids, utilizing standard regulatory sequences, selection markers, etc.). As many pharmaceutically important screens require human or model mammalian cell targets, retroviral vectors capable of transfecting such targets are preferred.

In a preferred embodiment, the fusion nucleic acids are part of a retroviral particle which infects the cells. Generally, infection of the cells is straightforward with the application of the infection-enhancing reagent polybrene, which is a polycation that facilitates viral binding to the target cell. Infection can be optimized such that each cell generally expresses a single construct, using the ratio of virus particles to number of cells. Infection follows a Poisson distribution.

In a preferred embodiment, the fusion nucleic acids are introduced into cells using retroviral vectors. Currently, the most efficient gene transfer methodologies harness the capacity of engineered viruses, such as retroviruses, to bypass natural cellular barriers to exogenous nucleic acid uptake. The use of recombinant retroviruses was pioneered by Richard Mulligan and David Baltimore with the Psi-2 lines and analogous retrovirus packaging systems, based on NIH 3T3 cells (see Mann et al., Cell 33:153-159 (1993), hereby incorporated by reference). Such helper-defective packaging lines are capable of producing all the necessary trans proteins -gag, pol, and env- that are required for packaging, processing, reverse transcription, and integration of recombinant genomes. Those RNA molecules that have in cis the ψ packaging signal are packaged into maturing virions. Retroviruses are preferred for a number of reasons. First, their derivation is easy. Second, unlike Adenovirus-mediated gene delivery, expression from retroviruses is long-term (adenoviruses do not integrate). Adeno-associated viruses have limited space for genes and regulatory units and there is some controversy as to their ability to integrate. Retroviruses therefore offer the best current compromise in terms of long-term expression, genomic flexibility, and stable integration, among other features. The main advantage of retroviruses is that their integration into the host genome allows for their stable transmission through cell division. This ensures that in cell types which undergo multiple independent maturation steps, such as hematopoietic cell progression, the retrovirus construct will remain resident and continue to express.

A particularly well suited retroviral transfection system is described in Mann et al., supra: Pear et al., PNAS USA 90(18):8392-6 (1993); Kitamura et al., PNAS USA 92:9146-9150 (1995); Kinsella et al., Human Gene Therapy 7:1405-1413; Hofmann et al., PNAS USA 93:5185-5190; Choate et al., Human Gene Therapy 7:2247 (1996); and WO 94/19478; and references cited therein, all of which are incorporated by reference.

In one embodiment of the invention, the library is generated in a intein-catalyzed cyclization scaffold. By "intein-catalyzed cyclization scaffold" herein is meant that the intein is engineered such that a cyclic peptide is generated upon intein-mediated splicing of the extein-intein junction points. Preferably, an intein cyclization scaffold includes the C-terminal intein motif, a library insert of 3 up to 30 amino acids in length, and the N-terminal intein motif. The C- and N-terminal intein motifs can be derived from any number of known inteins capable mediating protein splicing, including split-inteins.

Most wild-type inteins have requirements for a specific extein-encoded amino acid at the C-intein (IntB)/C-extein junction point. This varies depending on the intein, but most often consists of an cysteine, threonine or serine. Intein-generated cyclic peptide libraries may be generated in which this particular amino acid is fixed and corresponds to the amino acid present in the wild-type sequence. For example, the Ssp. DnaB intein utilizes an extein-encoded serine in this position.

A number of inteins have the ability to catalyze protein splicing when non-native amino acids are substituted at the C-intein (IntB)/C-extein junction point position. Degeneracy at the C-intein (IntB)/C-extein junction point position leads to cyclic peptide libraries of greater complexity and therefore added utility. The proposed degeneracy in this position most likely consists of a cysteine, serine or threonine but is not limited to these amino acids. The ability of a given intein-catalyzed cyclization scaffold to tolerate degeneracy at this position depends on the specific intein utilized and it's mechanism of protein splicing. Thus, isolation of intein cyclization scaffolds with a greater tolerance for degeneracy at the C-intein (IntB)/C-extein junction point is within the scope of this invention.

In one embodiment of the invention, the library is generated in a retrovirus DNA construct backbone, as is generally described in U.S.S.N. 08/789,333, filed January 23, 1997, incorporated herein by reference. Standard oligonucleotide synthesis is done to generate the random portion of the candidate bioactive agent, using techniques well known in the art (see Eckstein, Oligonucleotides and Analogues, A Practical Approach, IRL Press at Oxford University Press, 1991); libraries may be commercially purchased. Libraries with up to 10^9 to 10^{10} unique sequences can be readily generated in such DNA backbones. After generation of the DNA library, the library is cloned into a first primer. The first primer serves as a "cassette", which is inserted into the retroviral construct. The first primer generally contains a number of elements, including for example, the required regulatory sequences (e.g. translation, transcription, promoters, etc), fusion partners, restriction endonuclease (cloning and subcloning) sites, stop codons (preferably in all three frames), regions of complementarity for second strand priming (preferably at the end of the stop codon region as minor deletions or insertions may occur in the random region), etc. See U.S.S.N. 08/789,333, hereby incorporated by reference.

A second primer is then added, which generally consists of some or all of the complementarity region to prime the first primer and optional necessary sequences for a second unique restriction site for subcloning. DNA polymerase is added to make double-stranded oligonucleotides. The double-stranded oligonucleotides are cleaved with the appropriate subcloning restriction endonucleases and subcloned into the target retroviral vectors, described below.

Any number of suitable retroviral vectors may be used. Generally, the retroviral vectors may include: selectable marker genes under the control of internal ribosome entry sites (IRES), which allows for

bicistronic operons and thus greatly facilitates the selection of cells expressing peptides at uniformly high levels; and promoters driving expression of a second gene, placed in sense or anti-sense relative to the 5' LTR. Suitable selection genes include, but are not limited to, neomycin, blastocidin, bleomycin, puromycin, and hygromycin resistance genes, as well as self-fluorescent markers such as green fluorescent protein, enzymatic markers such as lacZ, and surface proteins such as CD8, etc.

Preferred vectors include a vector based on the murine stem cell virus (MSCV) (see Hawley et al., Gene Therapy 1:136 (1994)) and a modified MFG virus (Rivere et al., Genetics 92:6733 (1995)), and pBABE, outlined in the examples. A general schematic of the retroviral construct is depicted in Figure 6.

The retroviruses may include inducible and constitutive promoters. For example, there are situations wherein it is necessary to induce peptide expression only during certain phases of the selection process. For instance, a scheme to provide pro-inflammatory cytokines in certain instances must include induced expression of the peptides. This is because there is some expectation that over-expressed pro-inflammatory drugs might in the long-term be detrimental to cell growth. Accordingly, constitutive expression is undesirable, and the peptide is only turned on during that phase of the selection process when the phenotype is required, and then shut the peptide down by turning off the retroviral expression to confirm the effect or ensure long-term survival of the producer cells. A large number of both inducible and constitutive promoters are known.

In addition, it is possible to configure a retroviral vector to allow inducible expression of retroviral inserts after integration of a single vector in target cells; importantly, the entire system is contained within the single retrovirus. Tet-inducible retroviruses have been designed incorporating the Self-Inactivating (SIN) feature of 3' LTR enhancer/promoter retroviral deletion mutant (Hoffman et al., PNAS USA 93:5185 (1996)). Expression of this vector in cells is virtually undetectable in the presence of tetracycline or other active analogs. However, in the absence of Tet, expression is turned on to maximum within 48 hours after induction, with uniform increased expression of the whole population of cells that harbor the inducible retrovirus, indicating that expression is regulated uniformly within the infected cell population. A similar, related system uses a mutated Tet DNA-binding domain such that it bound DNA in the presence of Tet, and was removed in the absence of Tet. Either of these systems is suitable.

In this manner the primers create a library of fragments, each containing a different random nucleotide sequence that may encode a different peptide. The ligation products are then transformed into bacteria, such as *E. coli*, and DNA is prepared from the resulting library, as is generally outlined in Kitamura, PNAS USA 92:9146-9150 (1995), hereby expressly incorporated by reference.

Delivery of the library DNA into a retroviral packaging system results in conversion to infectious virus. Suitable retroviral packaging system cell lines include, but are not limited to, the Bing and BOSC23 cell lines described in WO 94/19478; Soneoka et al., Nucleic Acid Res. 23(4):628 (1995); Finer et al., Blood 83:43 (1994); Pheonix packaging lines such as PhiNX-eco and PhiNX-ampho, described below; 292T + gag-pol and retrovirus envelope; PA317; and cell lines outlined in Markowitz et al., Virology 167:400 (1988), Markowitz et al., J. Virol. 62:1120 (1988), Li et al., PNAS USA 93:11658 (1996), Kinsella et al., Human Gene Therapy 7:1405 (1996), all of which are incorporated by reference.

Preferred systems include PHEONIX-ECO and PHEONIX-AMPHO. Both PHEONIX-ECO and PHEONIX-AMPHO were tested for helper virus production and established as being helper-virus free. Both lines can carry episomes for the creation of stable cell lines which can be used to produce retrovirus. Both lines are readily testable by flow cytometry for stability of gag-pol (CD8) and envelope expression; after several months of testing the lines appear stable, and do not demonstrate loss of titre as did the first-generation lines BOSC23 and Bing (partly due to the choice of promoters driving expression of gag-pol and envelope). Both lines can also be used to transiently produce virus in a few days. Thus, these lines are fully compatible with transient, episomal stable, and library generation for retroviral gene transfer experiments. Finally, the titres produced by these lines have been tested. Using standard polybrene-enhanced retroviral infection, titres approaching or above 10^7 per ml were observed for both PHEONIX-eco and PHEONIX-ampho when carrying episomal constructs. When transiently produced virus is made, titres are usually 1/2 to 1/3 that value.

These lines are helper-virus free, carry episomes for long-term stable production of retrovirus, stably produce gag-pol and env, and do not demonstrate loss of viral titre over time. In addition, PhiNX-eco and PhiNX-ampho are capable of producing titres approaching or above 10^7 per ml when carrying episomal constructs, which, with concentration of virus, can be enhanced to 10^8 to 10^9 per ml.

In a preferred embodiment, the cell lines disclosed above, and the other methods for producing retrovirus, are useful for production of virus by transient transfection. The virus can either be used directly or be used to infect another retroviral producer cell line for "expansion" of the library.

Concentration of virus may be done as follows. Generally, retroviruses are titred by applying retrovirus-containing supernatant onto indicator cells, such as NIH3T3 cells, and then measuring the percentage of cells expressing phenotypic consequences of infection. The concentration of the virus is determined by multiplying the percentage of cells infected by the dilution factor involved, and taking into account the number of target cells available to obtain a relative titre. If the retrovirus contains a reporter gene, such as lacZ, then infection, integration, and expression of the recombinant virus is measured by histological staining for lacZ expression or by flow cytometry (FACS). In general,

retroviral titres generated from even the best of the producer cells do not exceed 10^7 per ml, unless concentration by relatively expensive or exotic apparatus. However, as it has been recently postulated that since a particle as large as a retrovirus will not move very far by brownian motion in liquid, fluid dynamics predicts that much of the virus never comes in contact with the cells to initiate the infection process. However, if cells are grown or placed on a porous filter and retrovirus is allowed to move past cells by gradual gravitometric flow, a high concentration of virus around cells can be effectively maintained at all times. Thus, up to a ten-fold higher infectivity by infecting cells on a porous membrane and allowing retrovirus supernatant to flow past them has been seen. This should allow titres of 10^9 after concentration.

The fusion nucleic acids and polypeptides of the invention are used to make cyclic peptides. By "cyclic peptides" or grammatical equivalents herein is meant the intracellular catalysis of peptide backbone cyclization. Generally, backbone cyclization results in the joining of the N and C termini of a peptide together such that a cyclic product is generated inside cells.

Preferably, every member of a peptide library is tested for bioactivity using one of the assays described below. For example, a cyclic peptide with 7 random positions has a complexity of $20^7 = 1.28 \times 10^9$, all of which will be tested for biological activity.

In the event it is not possible to test every member of a library for bioactivity, the library may be deliberately biased. For example, a cyclic peptide can be biased to cellular entry by fixing one or more relatively hydrophobic amino acids, such as tyrosine or tryptophan. Other types of biased libraries which may be synthesized include libraries which primary contain cyclic peptides comprising amino acids with large side chains and libraries in which the number of cyclic peptide conformers is restricted.

Highly restrained cyclic peptide libraries are made by using codons which code mainly for amino acids with large side chains. That is, when several resides of a cyclic peptide encode amino acids with large side chains, the conformation space of the peptide is restricted. The result is to bias the peptide to a higher affinity by reducing peptide conformational entropy. For example, a library of cyclic peptides could be created by restricting the triplet nucleotides coding for each random amino acid in the library to Cor T for the first position of the triplet, A, G or T for the second position in the triplet, and G, C or T for the third position in the triplet. This would result in a library biased to large amino acids, i.e., phenylalanine (F), leucine (L), tyrosine (Y), histidine (H), glutamine (Q), cysteine (C), tryptophan (W) and arginine (R). A library biased toward large amino acid side chains, combined with the loss of glycine, alanine, serine, threonine, aspartate, and glutamate results in a library coding for more rigid

peptides. As this library lacks an acidic amino acid, a pre-synthesized triplet coding glutamate (i.e., GAG) or aspartate (GAC) may be added during the DNA synthesis of the library.

Alternatively, a large amino acid side chain (i.e.) residue library may be created by pre-synthesizing triplets for desired residues. These residues are then mixed together during the DNA synthesis of the library. An example of a pre-synthesized large residue library is a library coding tyrosine (Y), arginine (R), glutamic acid (E), histidine (H), leucine (L), glutamine (Q), and optionally proline (P) or threonine (T).

A biased library can be created by restricting the number of conformers in a cyclic peptide. This approach is useful for structure activity relationship optimization. The number of conformers may be restricted by fixing a proline in the cyclic peptide ring at one position and leaving all of the other residues random. A smaller number of conformers allows for higher affinity binding interactions with target molecules, and more selective interactions with target moleucles due to a diminution of the possibility of "induced fit" binding. "Induced fit" comes at the expense of binding affinity due to a loss upon binding of the higher conformational entropy of a multi-conformer peptide. Higher affinity and selectivity are desirable for the development of cyclic peptides drugs. This is achieved by reducing the conformational entropy by including a rigid amino acid in a fixed position in each library member. For example, fixing one proline in a 7mer peptide is sufficient to restrict the conformational space of the cyclic peptide. For 8 to 10 mers, two prolines may be fixed in the ring allowing a diversity of $(20)^6$ or 6.4×10^7 in the 6 unfixed position of a 10 mer ring. Such a library is large enough to give hits in most screens for candidate drugs (as described below).

As will be appreciated by those in the art, the type of cells used in the present invention can vary widely. Basically, any mammalian cells may be used, with mouse, rat, primate and human cells being particularly preferred, although as will be appreciated by those in the art, modifications of the system by pseudotyping allows all eukaryotic cells to be used, preferably higher eukaryotes. As is more fully described below, a screen will be set up such that the cells exhibit a selectable phenotype in the presence of a cyclic peptide. As is more fully described below, cell types implicated in a wide variety of disease conditions are particularly useful, so long as a suitable screen may be designed to allow the selection of cells that exhibit an altered phenotype as a consequence of the presence of a cyclic peptide within the cell.

Accordingly, suitable cell types include, but are not limited to, tumor cells of all types (particularly melanoma, myeloid leukemia, carcinomas of the lung, breast, ovaries, colon, kidney, prostate, pancreas and testes), cardiomyocytes, endothelial cells, epithelial cells, lymphocytes (T-cell and B cell) , mast cells, eosinophils, vascular intimal cells, hepatocytes, leukocytes including mononuclear

leukocytes, stem cells such as haemopoetic, neural, skin, lung, kidney, liver and myocyte stem cells (for use in screening for differentiation and de-differentiation factors), osteoclasts, chondrocytes and other connective tissue cells, keratinocytes, melanocytes, liver cells, kidney cells, and adipocytes. Suitable cells also include known research cells, including, but not limited to, Jurkat T cells, NIH3T3 cells, CHO, Cos, etc. See the ATCC cell line catalog, hereby expressly incorporated by reference.

In one embodiment, the cells may be genetically engineered, that is, contain exogenous nucleic acid, for example, to contain target molecules.

Once made, the compositions of the invention find use in a number of applications. In particular, compositions with altered cyclization efficiency are made. The compositions of the invention also may be used to: (1) alter cellular phenotypes and/or physiology; (2) used in screening assays to identify target molecules associated with changes in cellular phenotype or phyisology; and, (3) used as drugs to treat a number of disease states, such as cancer, cardiovascular diseases, obesity, neurological disorders, etc.

In a preferred embodiment, inteins with altered cyclization activity are generated. Naturally occurring inteins are mutagenized and tested *in vivo* to determine whether the modified intein can catalyze protein or peptide cyclization in mammalian cells. Preferably, inteins so modified are characterized by more efficient cyclization kinetics *in vivo* or by the expression level of intein catalyzed cyclization scaffolds. Additional rounds of mutagenesis may be done to optimize *in vivo* function. Assays useful for measuring intein-catalyzed cyclization efficiency include fluorescent or gel based assays directly measuring cyclic peptide or protein levels, and functional assays based on the production of a functional cyclic peptide whose effects can be measured or selected for.

In a preferred embodiment, random mutagenesis (i.e. M13 primer mutagenesis and PCR mutagenesis), PCR shuffling or other directed evolution techniques are directed to a target codon or region and the resulting intein variants screened for altered cyclization activity. These techniques are well known and can be directed to predetermined sites, i.e., intein open reading frame or more specific regions or codons within.

Amino acid substitutions are typically of single residues; insertions usually will be on the order of from about 1 to 20 amino acids, although considerably larger insertions may be tolerated. Deletions range from about 1 to about 20 residues, although in some cases deletions may be much larger.

Substitutions, deletions, insertions or any combination thereof may be used to arrive at a final derivative. Generally these changes are done on a few amino acids to minimize the alteration of the

molecule. However, larger changes may be tolerated in certain circumstances. When small alterations in the characteristics of the intein protein are desired, substitutions are generally made in accordance with the following chart:

Chart I

Original Residue	Exemplary Substitutions
Ala	Ser
Arg	Lys
Asn	Gln, His
Asp	Glu
Cys	Ser
Gln	Asn
Glu	Asp
Gly	Pro
His	Asn, Gln
Ile	Leu, Val
Leu	Ile, Val
Lys	Arg, Gln, Glu
Met	Leu, Ile
Phe	Met, Leu, Tyr
Ser	Thr
Thr	Ser
Trp	Tyr
Tyr	Trp, Phe
Val	Ile, Leu

Substantial changes in function are made by selecting substitutions that are less conservative than those shown in Chart I. For example, substitutions may be made which more significantly affect: the structure of the polypeptide backbone in the area of the alteration, for example the alpha-helical or beta-sheet structure; the charge or hydrophobicity of the molecule at the target site; or the bulk of the side chain. The substitutions which in general are expected to produce the greatest changes in the polypeptide's properties are those in which (a) a hydrophilic residue, e.g. seryl or threonyl, is substituted for (or by) a hydrophobic residue, e.g. leucyl, isoleucyl, phenylalanyl, valyl or alanyl; (b) a cysteine or proline is substituted for (or by) any other residue; (c) a residue having an electropositive side chain, e.g. lysyl, arginyl, or histidyl, is substituted for (or by) an electronegative residue, e.g. glutamyl or aspartyl; or (d) a residue having a bulky side chain, e.g. phenylalanine, is substituted for (or by) one not having a side chain, e.g. glycine.

As outlined above, the variants typically exhibit the same qualitative biological activity (i.e. cyclization) although variants may be selected to modify other characteristics of the intein protein as needed. For example, endoplasmic reticulum/golgi directed intein libraries may be designed to operate in cellular environments more acidic than the cytoplasmic compartment.

In a preferred embodiment specific residues of an intein motif are substituted, resulting in proteins with modified characteristics. Such substitutions may occur at one or more residues, with 1-10 substitutions being preferred. Preferred characteristics to be modified include cyclization efficiency, half-life, stability, temperature sensitivity.

In a preferred embodiment, intein mutants are generated using PCR mutagenesis. The resulting mutants are screened for altered cyclization activity. By "altered" cyclization activity" refers to any characteristic or attribute of an intein that can be selected or detected and compared to the corresponding property of a naturally occurring intein. These properties include cyclization efficiency, stability, etc. Cyclization efficiency may be affected by the presence or absence of a given amino acid, the size of the peptide library, etc.

Unless otherwise specified, altered" cyclization activity, when comparing the cyclization efficiency of a mutant intein to the cyclization efficiency of wild-type or naturally occurring intein is preferably at least 1-fold, more preferably at least a 10-fold increase in activity.

Screens for mutants with improved cyclization efficiency can be done in procaryotes or eucaryotes. The mutants may be screened directly by assaying for the production of a cyclic peptide or indirectly by assaying a cyclic peptide's effects on a cell. Alternatively, the mutants may be screened indirectly by assaying the product of the cyclic peptide protein *in vitro*, e.g., enzyme inhibition assays, etc.

If the mutation prevents self-excision, no fluorescence is detected due to the interruption in the tertiary structure of GFP. If the mutation does not effect self-excision or enhances splicing efficiency, the degree of fluorescence may be quantified using a FACS analysis or other techniques known in the art. In addition, cyclization of the GFP reconstitutes the myc epitope which can be detected using Western analysis. T

In a preferred embodiment, a first plurality of cells is screened. That is, the cells into which the fusion nucleic acids are introduced are screened for an altered phenotype. Thus, in this embodiment, the effect of the bioactive peptide is seen in the same cells in which it is made; i.e. an autocrine effect.

By a "plurality of cells" herein is meant roughly from about 10^3 cells to 10^8 or 10^9, with from 10^6 to 10^8 being preferred. This plurality of cells comprises a cellular library, wherein generally each cell within the library contains a member of the peptide molecular library, i.e. a different peptide (or nucleic acid encoding the peptide), although as will be appreciated by those in the art, some cells within the library may not contain a peptide, and some may contain more than species of peptide. When methods other than retroviral infection are used to introduce the candidate nucleic acids into a plurality of cells, the

distribution of candidate nucleic acids within the individual cell members of the cellular library may vary widely, as it is generally difficult to control the number of nucleic acids which enter a cell during electroporation, etc.

In a preferred embodiment, the fusion nucleic acids are introduced into a first plurality of cells, and the effect of the peptide is screened in a second or third plurality of cells, different from the first plurality of cells, i.e. generally a different cell type. That is, the effect of the bioactive peptide is due to an extracellular effect on a second cell; i.e. an endocrine or paracrine effect. This is done using standard techniques. The first plurality of cells may be grown in or on one media, and the media is allowed to touch a second plurality of cells, and the effect measured. Alternatively, there may be direct contact between the cells. Thus, "contacting" is functional contact, and includes both direct and indirect. In this embodiment, the first plurality of cells may or may not be screened.

If necessary, the cells are treated to conditions suitable for the expression of the peptide (for example, when inducible promoters are used).

Thus, the methods of the present invention comprise introducing a molecular library of fusion nucleic acids encoding randomized peptides fused to scaffold into a plurality of cells, a cellular library. Each of the nucleic acids comprises a different nucleotide sequence encoding scaffold with a random peptide. The plurality of cells is then screened, as is more fully outlined below, for a cell exhibiting an altered phenotype. The altered phenotype is due to the presence of a bioactive peptide.

By "altered phenotype" or "changed physiology" or other grammatical equivalents herein is meant that the phenotype of the cell is altered in some way, preferably in some detectable and/or measurable way. As will be appreciated in the art, a strength of the present invention is the wide variety of cell types and potential phenotypic changes which may be tested using the present methods. Accordingly, any phenotypic change which may be observed, detected, or measured may be the basis of the screening methods herein. Suitable phenotypic changes include, but are not limited to: gross physical changes such as changes in cell morphology, cell growth, cell viability, adhesion to substrates or other cells, and cellular density; changes in the expression of one or more RNAs, proteins, lipids, hormones, cytokines, or other molecules; changes in the equilibrium state (i.e. half-life) or one or more RNAs, proteins, lipids, hormones, cytokines, or other molecules; changes in the localization of one or more RNAs, proteins, lipids, hormones, cytokines, or other molecules; changes in the bioactivity or specific activity of one or more RNAs, proteins, lipids, hormones, cytokines, receptors, or other molecules; changes in the secretion of ions, cytokines, hormones, growth factors, or other molecules; alterations in cellular membrane potentials, polarization, integrity or transport; changes in infectivity, susceptability, latency, adhesion, and uptake of viruses and bacterial pathogens; etc. By "capable of

altering the phenotype" herein is meant that the bioactive peptide can change the phenotype of the cell in some detectable and/or measurable way.

The altered phenotype may be detected in a wide variety of ways, as is described more fully below, and will generally depend and correspond to the phenotype that is being changed. Generally, the changed phenotype is detected using, for example: microscopic analysis of cell morphology; standard cell viability assays, including both increased cell death and increased cell viability, for example, cells that are now resistant to cell death via virus, bacteria, or bacterial or synthetic toxins; standard labeling assays such as fluorometric indicator assays for the presence or level of a particular cell or molecule, including FACS or other dye staining techniques; biochemical detection of the expression of target compounds after killing the cells; etc. In some cases, as is more fully described herein, the altered phenotype is detected in the cell in which the fusion nucleic acid was introduced; in other embodiments, the altered phenotype is detected in a second cell which is responding to some molecular signal from the first cell.

An altered phenotype of a cell indicates the presence of a bioactive peptide, acting preferably in a transdominant way. By "transdominant" herein is meant that the bioactive peptide indirectly causes the altered phenotype by acting on a second molecule, which leads to an altered phenotype. That is, a transdominant expression product has an effect that is not in cis, i.e., a trans event as defined in genetic terms or biochemical terms. A transdominant effect is a distinguishable effect by a molecular entity (i.e., the encoded peptide or RNA) upon some separate and distinguishable target; that is, not an effect upon the encoded entity itself. As such, transdominant effects include many well-known effects by pharmacologic agents upon target molecules or pathways in cells or physiologic systems; for instance, the β-lactam antibiotics have a transdominant effect upon peptidoglycan synthesis in bacterial cells by binding to penicillin binding proteins and disrupting their functions. An exemplary transdominant effect by a peptide is the ability to inhibit NF-κB signaling by binding to IκB-α at a region critical for its function, such that in the presence of sufficient amounts of the peptide (or molecular entity), the signaling pathways that normally lead to the activation of NF-κB through phosphorylation and/or degradation of IκB-α are inhibited from acting at IκB-α because of the binding of the peptide or molecular entity. In another instance, signaling pathways that are normally activated to secrete IgE are inhibited in the presence of peptide. Or, signaling pathways in adipose tissue cells, normally quiescent, are activated to metabolize fat. Or, in the presence of a peptide, intracellular mechanisms for the replication of certain viruses, such as HIV-I, or Herpes viridae family members, or Respiratory Syncytia Virus, for example, are inhibited.

A transdominant effect upon a protein or molecular pathway is clearly distinguishable from randomization, change, or mutation of a sequence within a protein or molecule of known or unknown

function to enhance or diminish a biochemical ability that protein or molecule already manifests. For instance, a protein that enzymatically cleaves β-lactam antibiotics, a β-lactamase, could be enhanced or diminished in its activity by mutating sequences internal to its structure that enhance or diminish the ability of this enzyme to act upon and cleave β-lactam antibiotics. This would be called a cis mutation to the protein. The effect of this protein upon β-lactam antibiotics is an activity the protein already manifests, to a distinguishable degree. Similarly, a mutation in the leader sequence that enhanced the export of this protein to the extracellular spaces wherein it might encounter β-lactam molecules more readily, or a mutation within the sequence that enhance the stability of the protein, would be termed cis mutations in the protein. For comparison, a transdominant effector of this protein would include an agent, independent of the β-lactamase, that bound to the β-lactamase in such a way that it enhanced or diminished the function of the β-lactamase by virtue of its binding to β-lactamase.

In a preferred embodiment, once a cell with an altered phenotype is detected, the presence of the fusion protein is verified, to ensure that the peptide was expressed and thus that the altered phenotype can be due to the presence of the peptide. As will be appreciated by those in the art, this verification of the presence of the peptide can be done either before, during or after the screening for an altered phenotype. This can be done in a variety of ways, although preferred methods utilize FACS techniques.

In a preferred embodiment, the devices of the invention comprise liquid handling components, including components for loading and unloading fluids at each station or sets of stations. The liquid handling systems can include robotic systems comprising any number of components. In addition, any or all of the steps outlined herein may be automated; thus, for example, the systems may be completely or partially automated.

As will be appreciated by those in the art, there are a wide variety of components which can be used, including, but not limited to, one or more robotic arms; plate handlers for the positioning of microplates; holders with cartridges and/or caps; automated lid or cap handlers to remove and replace lids for wells on non-cross contamination plates; tip assemblies for sample distribution with disposable tips; washable tip assemblies for sample distribution; 96 well loading blocks; cooled reagent racks; microtitler plate pipette positions (optionally cooled); stacking towers for plates and tips; and computer systems.

Fully robotic or microfluidic systems include automated liquid-, particle-, cell- and organism-handling including high throughput pipetting to perform all steps of screening applications. This includes liquid, particle, cell, and organism manipulations such as aspiration, dispensing, mixing, diluting, washing, accurate volumetric transfers; retrieving, and discarding of pipet tips; and repetitive pipetting of

identical volumes for multiple deliveries from a single sample aspiration. These manipulations are cross-contamination-free liquid, particle, cell, and organism transfers. This instrument performs automated replication of microplate samples to filters, membranes, and/or daughter plates, high-density transfers, full-plate serial dilutions, and high capacity operation.

In a preferred embodiment, chemically derivatized particles, plates, cartridges, tubes, magnetic particles, or other solid phase matrix with specificity to the assay components are used. The binding surfaces of microplates, tubes or any solid phase matrices include non-polar surfaces, highly polar surfaces, modified dextran coating to promote covalent binding, antibody coating, affinity media to bind fusion proteins or peptides, surface-fixed proteins such as recombinant protein A or G, nucleotide resins or coatings, and other affinity matrix are useful in this invention.

In a preferred embodiment, platforms for multi-well plates, multi-tubes, holders, cartridges, minitubes, deep-well plates, microfuge tubes, cryovials, square well plates, filters, chips, optic fibers, beads, and other solid-phase matrices or platform with various volumes are accommodated on an upgradable modular platform for additional capacity. This modular platform includes a variable speed orbital shaker, and multi-position work decks for source samples, sample and reagent dilution, assay plates, sample and reagent reservoirs, pipette tips, and an active wash station.

In a preferred embodiment, thermocycler and thermoregulating systems are used for stabilizing the temperature of the heat exchangers such as controlled blocks or platforms to provide accurate temperature control of incubating samples from 4°C to 100°C; this is in addition to or in place of the station thermocontrollers.

In a preferred embodiment, interchangeable pipet heads (single or multi-channel) with single or multiple magnetic probes, affinity probes, or pipetters robotically manipulate the liquid, particles, cells, and organisms. Multi-well or multi-tube magnetic separators or platforms manipulate liquid, particles, cells, and organisms in single or multiple sample formats.

In some embodiments, for example when electronic detection is not done, the instrumentation will include a detector, which can be a wide variety of different detectors, depending on the labels and assay. In a preferred embodiment, useful detectors include a microscope(s) with multiple channels of fluorescence; plate readers to provide fluorescent, ultraviolet and visible spectrophotometric detection with single and dual wavelength endpoint and kinetics capability, fluroescence resonance energy transfer (FRET), luminescence, quenching, two-photon excitation, and intensity redistribution; CCD cameras to capture and transform data and images into quantifiable formats; and a computer workstation.

These instruments can fit in a sterile laminar flow or fume hood, or are enclosed, self-contained systems, for cell culture growth and transformation in multi-well plates or tubes and for hazardous operations. The living cells will be grown under controlled growth conditions, with controls for temperature, humidity, and gas for time series of the live cell assays. Automated transformation of cells and automated colony pickers will facilitate rapid screening of desired cells.

Flow cytometry or capillary electrophoresis formats can be used for individual capture of magnetic and other beads, particles, cells, and organisms.

The flexible hardware and software allow instrument adaptability for multiple applications. The software program modules allow creation, modification, and running of methods. The system diagnostic modules allow instrument alignment, correct connections, and motor operations. The customized tools, labware, and liquid, particle, cell and organism transfer patterns allow different applications to be performed. The database allows method and parameter storage. Robotic and computer interfaces allow communication between instruments.

In a preferred embodiment, the robotic apparatus includes a central processing unit which communicates with a memory and a set of input/output devices (e.g., keyboard, mouse, monitor, printer, etc.) through a bus. Again, as outlined below, this may be in addition to or in place of the CPU for the multiplexing devices of the invention. The general interaction between a central processing unit, a memory, input/output devices, and a bus is known in the art. Thus, a variety of different procedures, depending on the experiments to be run, are stored in the CPU memory.

These robotic fluid handling systems can utilize any number of different reagents, including buffers, reagents, samples, washes, assay components such as label probes, etc.

Once the presence of the fusion protein is verified, the cell with the altered phenotype is generally isolated from the plurality which do not have altered phenotypes. This may be done in any number of ways, as is known in the art, and will in some instances depend on the assay or screen. Suitable isolation techniques include, but are not limited to, FACS, lysis selection using complement, cell cloning, scanning by Fluorimager, expression of a “survival” protein, induced expression of a cell surface protein or other molecule that can be rendered fluorescent or taggable for physical isolation; expression of an enzyme that changes a non-fluorescent molecule to a fluoroscent one; overgrowth against a background of no or slow growth; death of cells and isolation of DNA or other cell vitality indicator dyes, etc.

In a preferred embodiment, the fusion nucleic acid and/or the bioactive peptide (i.e. the fusion protein) is isolated from the positive cell. This may be done in a number of ways. In a preferred embodiment, primers complementary to DNA regions common to the retroviral constructs, or to specific components of the library such as a rescue sequence, defined above, are used to "rescue" the unique random sequence. Alternatively, the fusion protein is isolated using a rescue sequence. Thus, for example, rescue sequences comprising epitope tags or purification sequences may be used to pull out the fusion protein using immunoprecipitation or affinity columns. In some instances, as is outlined below, this may also pull out the primary target molecule, if there is a sufficiently strong binding interaction between the bioactive peptide and the target molecule. Alternatively, the peptide may be detected using mass spectroscopy.

Once rescued, the sequence of the bioactive peptide and/or fusion nucleic acid is determined. This information can then be used in a number of ways.

In a preferred embodiment, the bioactive peptide is resynthesized and reintroduced into the target cells, to verify the effect. This may be done using retroviruses, or alternatively using fusions to the HIV-1 Tat protein, and analogs and related proteins, which allows very high uptake into target cells. See for example, Fawell et al., PNAS USA 91:664 (1994); Frankel et al., Cell 55:1189 (1988); Savion et al., J. Biol. Chem. 256:1149 (1981); Derossi et al., J. Biol. Chem. 269:10444 (1994); and Baldin et al., EMBO J. 9:1511 (1990), all of which are incorporated by reference.

In a preferred embodiment, the sequence of a bioactive peptide is used to generate more candidate peptides. For example, the sequence of the bioactive peptide may be the basis of a second round of (biased) randomization, to develop bioactive peptides with increased or altered activities. Alternatively, the second round of randomization may change the affinity of the bioactive peptide. Furthermore, it may be desirable to put the identified random region of the bioactive peptide into other presentation structures, or to alter the sequence of the constant region of the presentation structure, to alter the conformation/shape of the bioactive peptide. It may also be desirable to "walk" around a potential binding site, in a manner similar to the mutagenesis of a binding pocket, by keeping one end of the ligand region constant and randomizing the other end to shift the binding of the peptide around.

In a preferred embodiment, either the bioactive peptide or the bioactive nucleic acid encoding it is used to identify target molecules, i.e. the molecules with which the bioactive peptide interacts. As will be appreciated by those in the art, there may be primary target molecules, to which the bioactive peptide binds or acts upon directly, and there may be secondary target molecules, which are part of the signalling pathway affected by the bioactive peptide; these might be termed "validated targets".

In a preferred embodiment, the bioactive peptide is a drug. As will be appreciated by those in the art, the structure of the cyclic peptide may be modeled and used in rational drug design to synthesize agents that mimic the interaction of the cyclic peptide with its' target. Drugs may also be modeled based on the three dimensional structure of the peptide bound to its target. Drugs so modeled may have structures that are similar to or unrelated to the starting structure of the cyclic peptide or the cyclic peptide bound to its target. Alternatively, high throughput screens can be used to identify small molecules capable of competing with the cyclic peptide for its target.

In a preferred embodiment, the bioactive cyclic peptide may be used as the starting point for designing/synthesizing derivative molecules with similar or more favorable properties for use as a drug. For example, individual amino acids, specific chemical groups, etc., can be replaced and the derivative molecule tested for use as a drug. Both naturally occurring and synthetic amino acid analogs (see below for definition) can be introduced in to the derivative molecule to optimize properties such as binding, stability, pharmocokinectics. Preferably, the derivative molecule has one or more of the following properties: improved stability, higher binding affinity, improved specificity for the target, improved pharmocokinetics, i.e., absorption, distribution , resistance to degradation, etc.

In a preferred embodiment, the bioactive peptide is used to pull out target molecules. For example, as outlined herein, if the target molecules are proteins, the use of epitope tags, purification sequences, or affinity tags can allow the purification of primary target molecules via biochemical means (co-immunoprecipitation, affinity columns, etc.). Alternatively, the peptide, when expressed in bacteria and purified, can be used as a probe against a bacterial cDNA expression library made from mRNA of the target cell type. Or, peptides can be used as "bait" in either yeast or mammalian two or three hybrid systems. Such interaction cloning approaches have been very useful to isolate DNA-binding proteins and other interacting protein components. The peptide(s) can be combined with other pharmacologic activators to study the epistatic relationships of signal transduction pathways in question. It is also possible to synthetically prepare labeled peptide and use it to screen a cDNA library expressed in bacteriophage for those cDNAs which bind the peptide. Furthermore, it is also possible that one could use cDNA cloning via retroviral libraries to "complement" the effect induced by the peptide. In such a strategy, the peptide would be required to be stochiometrically titrating away some important factor for a specific signaling pathway. If this molecule or activity is replenished by over-expression of a cDNA from within a cDNA library, then one can clone the target. Similarly, cDNAs cloned by any of the above yeast or bacteriophage systems can be reintroduced to mammalian cells in this manner to confirm that they act to complement function in the system the peptide acts upon.

In a preferred embodiment, target molecules are identified by incorporating an affinity tagged amino acid residue into the sequence of the cyclic peptide. For example, incorporation of a cysteine allows

for the chemical conjugation of the cyclic peptide to a solid support matrix via a disulfide bond. In particular, target molecules that bind to functional cyclic peptides are isolated and identified using affinity tagged amino acids.

In a preferred embodiment, the cysteine contributed by the extein is uniquely alkylated with an affinity reagent as part of the synthesis of the peptide to allow affinity extraction and identification of target molecules using HPLC-mass spectrometry methods. Cysteine-alkylated cyclic peptide analogs are tested for function, and if functional, target molecules are affinity extracted using methods well known in the art. If the cysteine-alkylated peptide analogs are not functional, synthetic cyclic peptide analogs are constructed with cysteine-affinity tag amino acid analogs in other positions and tested for function. In alternative embodiments, lysine affinity tagged amino acids are used.

Alternatively, if an affinity tagged amino acid cannot be produced *in vivo*, the tag can be introduced *in vitro* and tested *in vivo* for function.

Any amino acid which can be used as a affinity tag may be used in the methods of the invention. This includes both naturally occurring and synthetic amino acid analogs which can be introduced into the cyclic peptide to facilitate chemical conjugation or binding to a solid support matrix. Thus "amino acid", or "peptide residue", as used herein means both naturally occurring and synthetic amino acids. For example, homo-phenylalanine, citrulline, and noreleucine are considered amino acids for the purposes of the invention. "Amino acid" also includes imino acid residues such as proline and hydroxyproline. In addition, any amino acid can be replaced by the same amino acid but of the opposite chirality. Thus, any amino acid naturally occurring in the L-configuration (which may also be referred to as the R or S, depending upon the structure of the chemical entity) may be replaced with an amino acid of the same chemical structural type, but of the opposite chirality, generally referred to as the D- amino acid but which can additionally be referred to as the R- or the S-, depending upon its composition and chemical configuration. Such derivatives have the property of greatly increased stability, and therefore are advantageous in the formulation of compounds which may have longer in vivo half lives, when administered by oral, intravenous, intramuscular, intraperitoneal, topical, rectal, intraocular, or other routes.

In the preferred embodiment, the amino acids are in the (S) or L-configuration. If non-naturally occurring side chains are used, non-amino acid substituents may be used, for example to prevent or retard in vivo degradations. Proteins including non-naturally occurring amino acids may be synthesized or in some cases, made recombinantly; see van Hest et al., FEBS Lett 428:(1-2) 68-70 May 22 1998 and Tang et al., Abstr. Pap Am. Chem. S218:U138-U138 Part 2 August 22, 1999, both of which are expressly incorporated by reference herein.

Aromatic amino acids may be replaced with D- or L-naphylalanine, D- or L-Phenylglycine, D- or L-2-thieneylalanine, D- or L-1-, 2-, 3- or 4-pyreneylalanine, D- or L-3-thieneylalanine, D- or L-(2-pyridinyl)-alanine, D- or L-(3-pyridinyl)-alanine, D- or L-(2-pyrazinyl)-alanine, D- or L-(4-isopropyl)-phenylglycine, D-(trifluoromethyl)-phenylglycine, D-(trifluoromethyl)-phenylalanine, D-p-fluorophenylalanine, D- or L-p-biphenylphenylalanine, D- or L-p-methoxybiphenylphenylalanine, D- or L-2-indole(alkyl)alanines, and D- or L-alkylainines where alkyl may be substituted or unsubstituted methyl, ethyl, propyl, hexyl, butyl, pentyl, isopropyl, iso-butyl, sec-isotyl, iso-pentyl, non-acidic amino acids, of C1-C20.

Acidic amino acids can be substituted with non-carboxylate amino acids while maintaining a negative charge, and derivatives or analogs thereof, such as the non-limiting examples of (phosphono)alanine, glycine, leucine, isoleucine, threonine, or serine; or sulfated (e.g., -SO.sub.3 H) threonine, serine, tyrosine.

Other substitutions may include unnatural hyroxylated amino acids may made by combining "alkyl" with any natural amino acid. The term "alkyl" as used herein refers to a branched or unbranched saturated hydrocarbon group of 1 to 24 carbon atoms, such as methyl, ethyl, n-propyl, isoptopyl, n-butyl, isobutyl, t-butyl, octyl, decyl, tetradecyl, hexadecyl, eicosyl, tetracisyl and the like. Alkyl includes heteroalkyl, with atoms of nitrogen, oxygen and sulfur. Preferred alkyl groups herein contain 1 to 12 carbon atoms. Basic amino acids may be substituted with alkyl groups at any position of the naturally occurring amino acids lysine, arginine, ornithine, citrulline, or (guanidino)-acetic acid, or other (guanidino)alkyl-acetic acids, where "alkyl" is define as above. Nitrile derivatives (e.g., containing the CN-moiety in place of COOH) may also be substituted for asparagine or glutamine, and methionine sulfoxide may be substituted for methionine. Methods of preparation of such peptide derivatives are well known to one skilled in the art.

In addition, any amide linkage can be replaced by a ketomethylene moiety. Such derivatives are expected to have the property of increased stability to degradation by enzymes, and therefore possess advantages for the formulation of compounds which may have increased in vivo half lives, as administered by oral, intravenous, intramuscular, intraperitoneal, topical, rectal, intraocular, or other routes.

Additional amino acid modifications of amino acids of to the present invention may include the following: Cysteinyl residues may be reacted with alpha-haloacetates (and corresponding amines), such as 2-chloroacetic acid or chloroacetamide, to give carboxymethyl or carboxyamidomethyl derivatives. Cysteinyl residues may also be derivatized by reaction with compounds such as bromotrifluoroacetone, alpha-bromo-beta-(5-imidozoyl)propionic acid, chloroacetyl phosphate, N-alkylmaleimides, 3-nitro-2-pyridyl disulfide, methyl 2-pyridyl disulfide, p-chloromercuribenzoate, 2-

chloromercuri-4-nitrophenol, or chloro-7-nitrobenzo-2-oxa-1,3-diazole.

Histidyl residues may be derivatized by reaction with compounds such as diethylprocarbonate e.g., at pH 5.5-7.0 because this agent is relatively specific for the histidyl side chain, and para-bromophenacyl bromide may also be used; e.g., where the reaction is preferably performed in 0.1M sodium cacodylate at pH 6.0.

Lysinyl and amino terminal residues may be reacted with compounds such as succinic or other carboxylic acid anhydrides. Derivatization with these agents is expected to have the effect of reversing the charge of the lysinyl residues. Other suitable reagents for derivatizing alpha-amino-containing residues include compounds such as imidoesters/e.g., as methyl picolinimidate; pyridoxal phosphate; pyridoxal; chloroborohydride; trinitrobenzenesulfonic acid; O-methylisourea; 2,4 pentanedione; and transaminase-catalyzed reaction with glyoxylate.

Arginyl residues may be modified by reaction with one or several conventional reagents, among them phenylglyoxal, 2,3-butanedione, 1,2-cyclohexanedione, and ninhydrin according to known method steps. Derivatization of arginine residues requires that the reaction be performed in alkaline conditions because of the high pKa of the guanidine functional group. Furthermore, these reagents may react with the groups of lysine as well as the arginine epsilon-amino group.

The specific modification of tyrosyl residues per se is well-known, such as for introducing spectral labels into tyrosyl residues by reaction with aromatic diazonium compounds or tetranitromethane. N-acetylimidizol and tetranitromethane may be used to form O-acetyl tyrosyl species and 3-nitro derivatives, respectively.

Carboxyl side groups (aspartyl or glutamyl) may be selectively modified by reaction with carbodiimides (R'-N-C-N-R') such as 1-cyclohexyl-3-(2-morpholinyl- (4-ethyl) carbodiimide or 1-ethyl-3-(4-azonia-4,4-dimethylpentyl) carbodiimide. Furthermore aspartyl and glutamyl residues may be converted to asparaginyl and glutaminyl residues by reaction with ammonium ions.

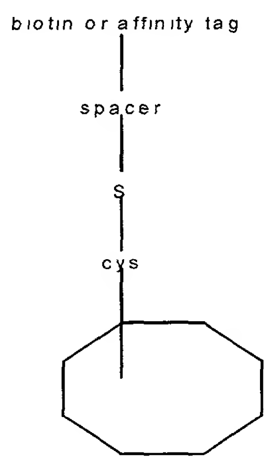
Glutaminyl and asparaginyl residues may be frequently deamidated to the corresponding glutamyl and aspartyl residues. Alternatively, these residues may be deamidated under mildly acidic conditions. Either form of these residues falls within the scope of the present invention.

Examples of affinity labeled amino acids useful for extraction of target molecules include lysine-epsilon amino biotin, or lysine reacted with amine-specific biotinylation reagents such as biotin-NHS ester and sulfo-NHS biotin.

Spacers may be incorporated between the affinity element and the peptide to relieve steric restraints between the affinity tag and a cyclic peptide bound to a target molecule. A spacer which may be used with affinity tagged lysine is NHS-LC-biotin (Pierce Checmical CO., Rockford IL), although other spacers as are known in the art also may be used.

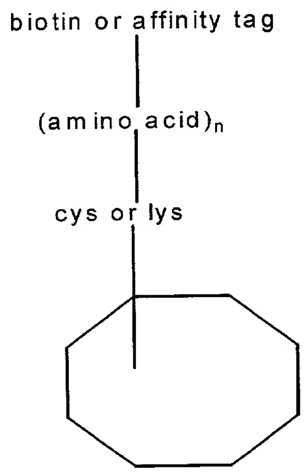
Examples of spacers which can be used with affinity tagged cysteines include cysteine reacted with iodoacetamido-biotin, biotin-hexyl-3'-(2'-pyridyldithio) propionamide (a 29 Å spacer from Pierce Chemical), iodoacetyl-LC-biotin (27 Å spacer) or biotin-BMCC with a 32 Å spacer (Pierce Chemical). An example of a spacer used with affinity tagged cysteine is shown in Structure 1:

Structure 1



Alternatively, as part of the solid phase synthesis of the peptide, affinity tags may be synthesized branching off from the cysteine or lysine. In this case, the spacer consists of a defined number (i.e. n) of amino acids branching off the side chain of the cysteine or lysine or another residue of the cyclic peptide. Preferably, n = 1 to 40. This allows for spacers of variable length, ranging from 3 Å to 100 Å or more. Gycines, because of their flexibility, are preferred because a sterically bulky target molecules bound to the cyclic peptide can be accommodated. The affinity tag is inserted at the end of the side chain as illustrated in Structure 2:

Structure 2



In a preferred embodiment, the spacer is at least one protein diameter long (20-40 Å). When the interacting target molecule is part of a large complex, the spacer is up to at least two protein diameters (40-80 Å).

Once primary target molecules have been identified, secondary target molecules may be identified in the same manner, using the primary target as the "bait". In this manner, signaling pathways may be elucidated. Similarly, bioactive peptides specific for secondary target molecules may also be discovered, to allow a number of bioactive peptides to act on a single pathway, for example for combination therapies.

The screening methods of the present invention may be useful to screen a large number of cell types under a wide variety of conditions. Generally, the host cells are cells that are involved in disease states, and they are tested or screened under conditions that normally result in undesirable consequences on the cells. When a suitable bioactive peptide is found, the undesirable effect may be reduced or eliminated. Alternatively, normally desirable consequences may be reduced or eliminated, with an eye towards elucidating the cellular mechanisms associated with the disease state or signalling pathway.

In a preferred embodiment, the present methods are useful in cancer applications. The ability to rapidly and specifically kill tumor cells is a cornerstone of cancer chemotherapy. In general, using the methods of the present invention, random libraries can be introduced into any tumor cell (primary or cultured), and peptides identified which by themselves induce apoptosis, cell death, loss of cell division or decreased cell growth. This may be done de novo, or by biased randomization toward

known peptide agents, such as angiostatin, which inhibits blood vessel wall growth. Alternatively, the methods of the present invention can be combined with other cancer therapeutics (e.g. drugs or radiation) to sensitize the cells and thus induce rapid and specific apoptosis, cell death, loss of cell division or decreased cell growth after exposure to a secondary agent. Similarly, the present methods may be used in conjunction with known cancer therapeutics to screen for agonists to make the therapeutic more effective or less toxic. This is particularly preferred when the chemotherapeutic is very expensive to produce such as taxol.

Known oncogenes such as v-Abl, v-Src, v-Ras, and others, induce a transformed phenotype leading to abnormal cell growth when transfected into certain cells. This is also a major problem with micro-metastases. Thus, in a preferred embodiment, non-transformed cells can be transfected with these oncogenes, and then random libraries introduced into these cells, to select for bioactive peptides which reverse or correct the transformed state. One of the signal features of oncogene transformation of cells is the loss of contact inhibition and the ability to grow in soft-agar. When transforming viruses are constructed containing v-Abl, v-Src, or v-Ras in IRES-puro retroviral vectors, infected into target 3T3 cells, and subjected to puromycin selection, all of the 3T3 cells hyper-transform and detach from the plate. The cells may be removed by washing with fresh medium. This can serve as the basis of a screen, since cells which express a bioactive peptide will remain attached to the plate and form colonies.

Similarly, the growth and/or spread of certain tumor types is enhanced by stimulatory responses from growth factors and cytokines (PDGF, EGF, Heregulin, and others) which bind to receptors on the surfaces of specific tumors. In a preferred embodiment, the methods of the invention are used to inhibit or stop tumor growth and/or spread, by finding bioactive peptides capable of blocking the ability of the growth factor or cytokine to stimulate the tumor cell. The introduction of random libraries into specific tumor cells with the addition of the growth factor or cytokine, followed by selection of bioactive peptides which block the binding, signaling, phenotypic and/or functional responses of these tumor cells to the growth factor or cytokine in question.

Similarly, the spread of cancer cells (invasion and metastasis) is a significant problem limiting the success of cancer therapies. The ability to inhibit the invasion and/or migration of specific tumor cells would be a significant advance in the therapy of cancer. Tumor cells known to have a high metastatic potential (for example, melanoma, lung cell carcinoma, breast and ovarian carcinoma) can have random libraries introduced into them, and peptides selected which in a migration or invasion assay, inhibit the migration and/or invasion of specific tumor cells. Particular applications for inhibition of the metastatic phenotype, which could allow a more specific inhibition of metastasis, include the metastasis suppressor gene NM23, which codes for a dinucleoside diphosphate kinase. Thus

intracellular peptide activators of this gene could block metastasis, and a screen for its upregulation (by fusing it to a reporter gene) would be of interest. Many oncogenes also enhance metastasis. Peptides which inactivate or counteract mutated RAS oncogenes, v-MOS, v-RAF, A-RAF, v-SRC, v-FES, and v-FMS would also act as anti-metastatics. Peptides which act intracellularly to block the release of combinations of proteases required for invasion, such as the matrix metalloproteases and urokinase, could also be effective antimetastatics.

In a preferred embodiment, the random libraries of the present invention are introduced into tumor cells known to have inactivated tumor suppressor genes, and successful reversal by either reactivation or compensation of the knockout would be screened by restoration of the normal phenotype. A major example is the reversal of p53-inactivating mutations, which are present in 50% or more of all cancers. Since p53's actions are complex and involve its action as a transcription factor, there are probably numerous potential ways a peptide or small molecule derived from a peptide could reverse the mutation. One example would be upregulation of the immediately downstream cyclin-dependent kinase p21CIP1/WAF1. To be useful such reversal would have to work for many of the different known p53 mutations. This is currently being approached by gene therapy; one or more small molecules which do this might be preferable.

Another example involves screening of bioactive peptides which restore the constitutive function of the brca-1 or brca-2 genes, and other tumor suppressor genes important in breast cancer such as the adenomatous polyposis coli gene (APC) and the Drosophila discs-large gene (Dlg), which are components of cell-cell junctions. Mutations of brca-1 are important in hereditary ovarian and breast cancers, and constitute an additional application of the present invention.

In a preferred embodiment, the methods of the present invention are used to create novel cell lines from cancers from patients. A retrovirally delivered short peptide which inhibits the final common pathway of programmed cell death should allow for short- and possibly long-term cell lines to be established. Conditions of *in vitro* culture and infection of human leukemia cells will be established. There is a real need for methods which allow the maintenance of certain tumor cells in culture long enough to allow for physiological and pharmacological studies. Currently, some human cell lines have been established by the use of transforming agents such as Epstein-Barr virus that considerably alters the existing physiology of the cell. On occasion, cells will grow on their own in culture but this is a random event. Programmed cell death (apoptosis) occurs via complex signaling pathways within cells that ultimately activate a final common pathway producing characteristic changes in the cell leading to a non-inflammatory destruction of the cell. It is well known that tumor cells have a high apoptotic index, or propensity to enter apoptosis *in vivo*. When cells are placed in culture, the *in vivo* stimuli for malignant cell growth are removed and cells readily undergo apoptosis. The objective would be to

develop the technology to establish cell lines from any number of primary tumor cells, for example primary human leukemia cells, in a reproducible manner without altering the native configuration of the signaling pathways in these cells. By introducing nucleic acids encoding peptides which inhibit apoptosis, increased cell survival *in vitro*, and hence the opportunity to study signalling transduction pathways in primary human tumor cells, is accomplished. In addition, these methods may be used for culturing primary cells, i.e. non-tumor cells.

In a preferred embodiment, the present methods are useful in cardiovascular applications. In a preferred embodiment, cardiomyocytes may be screened for the prevention of cell damage or death in the presence of normally injurious conditions, including, but not limited to, the presence of toxic drugs (particularly chemotherapeutic drugs), for example, to prevent heart failure following treatment with adriamycin; anoxia, for example in the setting of coronary artery occlusion; and autoimmune cellular damage by attack from activated lymphoid cells (for example as seen in post viral myocarditis and lupus). Candidate bioactive peptides are inserted into cardiomyocytes, the cells are subjected to the insult, and bioactive peptides are selected that prevent any or all of: apoptosis; membrane depolarization (i.e. decrease arrythmogenic potential of insult); cell swelling; or leakage of specific intracellular ions, second messengers and activating molecules (for example, arachidonic acid and/or lysophosphatidic acid).

In a preferred embodiment, the present methods are used to screen for diminished arrhythmia potential in cardiomyocytes. The screens comprise the introduction of the candidate nucleic acids encoding candidate bioactive peptides, followed by the application of arrythmogenic insults, with screening for bioactive peptides that block specific depolarization of cell membrane. This may be detected using patch clamps, or via fluorescence techniques). Similarly, channel activity (for example, potassium and chloride channels) in cardiomyocytes could be regulated using the present methods in order to enhance contractility and prevent or diminish arrhythmias.

In a preferred embodiment, the present methods are used to screen for enhanced contractile properties of cardiomyocytes and diminish heart failure potential. The introduction of the libraries of the invention followed by measuring the rate of change of myosin polymerization/depolymerization using fluorescent techniques can be done. Bioactive peptides which increase the rate of change of this phenomenon can result in a greater contractile response of the entire myocardium, similar to the effect seen with digitalis.

In a preferred embodiment, the present methods are useful to identify agents that will regulate the intracellular and sarcolemmal calcium cycling in cardiomyocytes in order to prevent arrhythmias.

Bioactive peptides are selected that regulate sodium-calcium exchange, sodium proton pump function, and regulation of calcium-ATPase activity.

In a preferred embodiment, the present methods are useful to identify agents that diminish embolic phenomena in arteries and arterioles leading to strokes (and other occlusive events leading to kidney failure and limb ischemia) and angina precipitating a myocardial infarct are selected. For example, bioactive peptides which will diminish the adhesion of platelets and leukocytes, and thus diminish the occlusion events. Adhesion in this setting can be inhibited by the libraries of the invention being inserted into endothelial cells (quiescent cells, or activated by cytokines, i.e. IL-1, and growth factors, i.e. PDGF / EGF) and then screening for peptides that either: 1) down regulate adhesion molecule expression on the surface of the endothelial cells (binding assay); 2) block adhesion molecule activation on the surface of these cells (signaling assay); or 3) release in an autocrine manner peptides that block receptor binding to the cognate receptor on the adhering cell.

Embolic phenomena can also be addressed by activating proteolytic enzymes on the cell surfaces of endothelial cells, and thus releasing active enzyme which can digest blood clots. Thus, delivery of the libraries of the invention to endothelial cells is done, followed by standard fluorogenic assays, which will allow monitoring of proteolytic activity on the cell surface towards a known substrate. Bioactive peptides can then be selected which activate specific enzymes towards specific substrates.

In a preferred embodiment, arterial inflammation in the setting of vasculitis and post-infarction can be regulated by decreasing the chemotactic responses of leukocytes and mononuclear leukocytes. This can be accomplished by blocking chemotactic receptors and their responding pathways on these cells. Candidate bioactive libraries can be inserted into these cells, and the chemotactic response to diverse chemokines (for example, to the IL-8 family of chemokines, RANTES) inhibited in cell migration assays.

In a preferred embodiment, arterial restenosis following coronary angioplasty can be controlled by regulating the proliferation of vascular intimal cells and capillary and/or arterial endothelial cells. Candidate bioactive peptide libraries can be inserted into these cell types and their proliferation in response to specific stimuli monitored. One application may be intracellular peptides which block the expression or function of c-myc and other oncogenes in smooth muscle cells to stop their proliferation. A second application may involve the expression of libraries in vascular smooth muscle cells to selectively induce their apoptosis. Application of small molecules derived from these peptides may require targeted drug delivery; this is available with stents, hydrogel coatings, and infusion-based catheter systems. Peptides which downregulate endothelin-1A receptors or which block the release of the potent vasoconstrictor and vascular smooth muscle cell mitogen endothelin-1 may also be

candidates for therapeutics. Peptides can be isolated from these libraries which inhibit growth of these cells, or which prevent the adhesion of other cells in the circulation known to release autocrine growth factors, such as platelets (PDGF) and mononuclear leukocytes.

The control of capillary and blood vessel growth is an important goal in order to promote increased blood flow to ischemic areas (growth), or to cut-off the blood supply (angiogenesis inhibition) of tumors. Candidate bioactive peptide libraries can be inserted into capillary endothelial cells and their growth monitored. Stimuli such as low oxygen tension and varying degrees of angiogenic factors can regulate the responses, and peptides isolated that produce the appropriate phenotype. Screening for antagonism of vascular endothelial cell growth factor, important in angiogenesis, would also be useful.

In a preferred embodiment, the present methods are useful in screening for decreases in atherosclerosis producing mechanisms to find peptides that regulate LDL and HDL metabolism. Candidate libraries can be inserted into the appropriate cells (including hepatocytes, mononuclear leukocytes, endothelial cells) and peptides selected which lead to a decreased release of LDL or diminished synthesis of LDL, or conversely to an increased release of HDL or enhanced synthesis of HDL. Bioactive peptides can also be isolated from candidate libraries which decrease the production of oxidized LDL, which has been implicated in atherosclerosis and isolated from atherosclerotic lesions. This could occur by decreasing its expression, activating reducing systems or enzymes, or blocking the activity or production of enzymes implicated in production of oxidized LDL, such as 15-lipoxygenase in macrophages.

In a preferred embodiment, the present methods are used in screens to regulate obesity via the control of food intake mechanisms or diminishing the responses of receptor signaling pathways that regulate metabolism. Bioactive peptides that regulate or inhibit the responses of neuropeptide Y (NPY), cholecystokinin and galanin receptors, are particularly desirable. Candidate libraries can be inserted into cells that have these receptors cloned into them, and inhibitory peptides selected that are secreted in an autocrine manner that block the signaling responses to galanin and NPY. In a similar manner, peptides can be found that regulate the leptin receptor.

In a preferred embodiment, the present methods are useful in neurobiology applications. Candidate libraries may be used for screening for anti-apoptotics for preservation of neuronal function and prevention of neuronal death. Initial screens would be done in cell culture. One application would include prevention of neuronal death, by apoptosis, in cerebral ischemia resulting from stroke. Apoptosis is known to be blocked by neuronal apoptosis inhibitory protein (NAIP); screens for its upregulation, or effecting any coupled step could yield peptides which selectively block neuronal

apoptosis. Other applications include neurodegenerative diseases such as Alzheimer's disease and Huntington's disease.

In a preferred embodiment, the present methods are useful in bone biology applications. Osteoclasts are known to play a key role in bone remodeling by breaking down "old" bone, so that osteoblasts can lay down "new" bone. In osteoporosis one has an imbalance of this process. Osteoclast overactivity can be regulated by inserting candidate libraries into these cells, and then looking for bioactive peptides that produce: 1) a diminished processing of collagen by these cells; 2) decreased pit formation on bone chips; and 3) decreased release of calcium from bone fragments.

The present methods may also be used to screen for agonists of bone morphogenic proteins, hormone mimetics to stimulate, regulate, or enhance new bone formation (in a manner similar to parathyroid hormone and calcitonin, for example). These have use in osteoporosis, for poorly healing fractures, and to accelerate the rate of healing of new fractures. Furthermore, cell lines of connective tissue origin can be treated with candidate libraries and screened for their growth, proliferation, collagen stimulating activity, and/or proline incorporating ability on the target osteoblasts. Alternatively, candidate libraries can be expressed directly in osteoblasts or chondrocytes and screened for increased production of collagen or bone.

In a preferred embodiment, the present methods are useful in skin biology applications. Keratinocyte responses to a variety of stimuli may result in psoriasis, a proliferative change in these cells. Candidate libraries can be inserted into cells removed from active psoriatic plaques, and bioactive peptides isolated which decrease the rate of growth of these cells.

In a preferred embodiment, the present methods are useful in the regulation or inhibition of keloid formation (i.e. excessive scarring). Candidate libraries inserted into skin connective tissue cells isolated from individuals with this condition, and bioactive peptides isolated that decrease proliferation, collagen formation, or proline incorporation. Results from this work can be extended to treat the excessive scarring that also occurs in burn patients. If a common peptide motif is found in the context of the keloid work, then it can be used widely in a topical manner to diminish scarring post burn.

Similarly, wound healing for diabetic ulcers and other chronic "failure to heal" conditions in the skin and extremities can be regulated by providing additional growth signals to cells which populate the skin and dermal layers. Growth factor mimetics may in fact be very useful for this condition. Candidate libraries can be inserted into skin connective tissue cells, and bioactive peptides isolated which promote the growth of these cells under "harsh" conditions, such as low oxygen tension, low pH, and the presence of inflammatory mediators.

Cosmeceutical applications of the present invention include the control of melanin production in skin melanocytes. A naturally occurring peptide, arbutin, is a tyrosine hydroxylase inhibitor, a key enzyme in the synthesis of melanin. Candidate libraries can be inserted into melanocytes and known stimuli that increase the synthesis of melanin applied to the cells. Bioactive peptides can be isolated that inhibit the synthesis of melanin under these conditions.

In a preferred embodiment, the present methods are useful in endocrinology applications. The retroviral peptide library technology can be applied broadly to any endocrine, growth factor, cytokine or chemokine network which involves a signaling peptide or protein that acts in either an endocrine, paracrine or autocrine manner that binds or dimerizes a receptor and activates a signaling cascade that results in a known phenotypic or functional outcome. The methods are applied so as to isolate a peptide which either mimics the desired hormone (i.e., insulin, leptin, calcitonin, PDGF, EGF, EPO, GMCSF, IL1-17, mimetics) or inhibits its action by either blocking the release of the hormone, blocking its binding to a specific receptor or carrier protein (for example, CRF binding protein), or inhibiting the intracellular responses of the specific target cells to that hormone. Selection of peptides which increase the expression or release of hormones from the cells which normally produce them could have broad applications to conditions of hormonal deficiency.

In a preferred embodiment, the present methods are useful in infectious disease applications. Viral latency (herpes viruses such as CMV, EBV, HBV, and other viruses such as HIV) and their reactivation are a significant problem, particularly in immunosuppressed patients (patients with AIDS and transplant patients). The ability to block the reactivation and spread of these viruses is an important goal. Cell lines known to harbor or be susceptible to latent viral infection can be infected with the specific virus, and then stimuli applied to these cells which have been shown to lead to reactivation and viral replication. This can be followed by measuring viral titers in the medium and scoring cells for phenotypic changes. Candidate libraries can then be inserted into these cells under the above conditions, and peptides isolated which block or diminish the growth and/or release of the virus. As with chemotherapeutics, these experiments can also be done with drugs which are only partially effective towards this outcome, and bioactive peptides isolated which enhance the virucidal effect of these drugs. Bioactive peptides may also be tested for the ability to block some aspect of viral assembly, viral replication, entry or infectious cycle.

One example of many is the ability to block HIV-1 infection. HIV-1 requires CD4 and a co-receptor which can be one of several seven transmembrane G-protein coupled receptors. In the case of the infection of macrophages, CCR-5 is the required co-receptor, and there is strong evidence that a block on CCR-5 will result in resistance to HIV-1 infection. There are two lines of evidence for this statement. First, it is known that the natural ligands for CCR-5, the CC chemokines RANTES, MIP1a

and MIP1b are responsible for CD8+ mediated resistance to HIV. Second, individuals homozygous for a mutant allele of CCR-5 are completely resistant to HIV infection. Thus, an inhibitor of the CCR-5/HIV interaction would be of enormous interest to both biologists and clinicians. The extracellular anchored constructs offer superb tools for such a discovery. Into the transmembrane, epitope tagged, glycine-serine tethered constructs (ssTM V G20 E TM), one can place a random, cyclized peptide library of the general sequence CNNNNNNNNNNC or C-$(X)_n$-C. Then one infects a cell line that expresses CCR-5 with retroviruses containing this library. Using an antibody to CCR-5 one can use FACS to sort desired cells based on the binding of this antibody to the receptor. All cells which do not bind the antibody will be assumed contain inhibitors of this antibody binding site. These inhibitors, in the retroviral construct can be further assayed for their ability to inhibit HIV-1 entry.

Viruses are known to enter cells using specific receptors to bind to cells (for example, HIV uses CD4, coronavirus uses CD13, murine leukemia virus uses transport protein, and measles virus usesCD44) and to fuse with cells (HIV uses chemokine receptor). Candidate libraries can be inserted into target cells known to be permissive to these viruses, and bioactive peptides isolated which block the ability of these viruses to bind and fuse with specific target cells.

Intein libraries may also be used to screen for cyclic peptides which block HIV-1 infection. For example, inteins can be designed such that cyclized peptides are secreted from cells where they can bind to CCR5 and antagonize HIV-1 binding.

In a preferred embodiment, the present invention finds use with infectious organisms. Intracellular organisms such as mycobacteria, listeria, salmonella, pneumocystis, yersinia, leishmania, T. cruzi, can persist and replicate within cells, and become active in immunosuppressed patients. There are currently drugs on the market and in development which are either only partially effective or ineffective against these organisms. Candidate libraries can be inserted into specific cells infected with these organisms (pre- or post-infection), and bioactive peptides selected which promote the intracellular destruction of these organisms in a manner analogous to intracellular "antibiotic peptides" similar to magainins. In addition peptides can be selected which enhance the cidal properties of drugs already under investigation which have insufficient potency by themselves, but when combined with a specific peptide from a candidate library, are dramatically more potent through a synergistic mechanism. Finally, bioactive peptides can be isolated which alter the metabolism of these intracellular organisms, in such a way as to terminate their intracellular life cycle by inhibiting a key organismal event.

Antibiotic drugs that are widely used have certain dose dependent, tissue specific toxicities. For example renal toxicity is seen with the use of gentamicin, tobramycin, and amphotericin; hepatotoxicity is seen with the use of INH and rifampin; bone marrow toxicity is seen with chloramphenicol; and

platelet toxicity is seen with ticarcillin, etc. These toxicities limit their use. Candidate libraries can be introduced into the specific cell types where specific changes leading to cellular damage or apoptosis by the antibiotics are produced, and bioactive peptides can be isolated that confer protection, when these cells are treated with these specific antibiotics.

Furthermore, the present invention finds use in screening for bioactive peptides that block antibiotic transport mechanisms. The rapid secretion from the blood stream of certain antibiotics limits their usefulness. For example penicillins are rapidly secreted by certain transport mechanisms in the kidney and choroid plexus in the brain. Probenecid is known to block this transport and increase serum and tissue levels. Candidate agents can be inserted into specific cells derived from kidney cells and cells of the choroid plexus known to have active transport mechanisms for antibiotics. Bioactive peptides can then be isolated which block the active transport of specific antibiotics and thus extend the serum halflife of these drugs.

In a preferred embodiment, the present methods are useful in drug toxicities and drug resistance applications. Drug toxicity is a significant clinical problem. This may manifest itself as specific tissue or cell damage with the result that the drug's effectiveness is limited. Examples include myeloablation in high dose cancer chemotherapy, damage to epithelial cells lining the airway and gut, and hair loss. Specific examples include adriamycin induced cardiomyocyte death, cisplatinin-induced kidney toxicity, vincristine-induced gut motility disorders, and cyclosporin-induced kidney damage. Candidate libraries can be introduced into specific cell types with characteristic drug-induced phenotypic or functional responses, in the presence of the drugs, and agents isolated which reverse or protect the specific cell type against the toxic changes when exposed to the drug. These effects may manifest as blocking the drug induced apoptosis of the cell of interest, thus initial screens will be for survival of the cells in the presence of high levels of drugs or combinations of drugs used in combination chemotherapy.

Drug toxicity may be due to a specific metabolite produced in the liver or kidney which is highly toxic to specific cells, or due to drug interactions in the liver which block or enhance the metabolism of an administered drug. Candidate libraries can be introduced into liver or kidney cells following the exposure of these cells to the drug known to produce the toxic metabolite. Bioactive peptides can be isolated which alter how the liver or kidney cells metabolize the drug, and specific agents identified which prevent the generation of a specific toxic metabolite. The generation of the metabolite can be followed by mass spectrometry, and phenotypic changes can be assessed by microscopy. Such a screen can also be done in cultured hepatocytes, cocultured with readout cells which are specifically sensitive to the toxic metabolite. Applications include reversible (to limit toxicity) inhibitors of enzymes involved in drug metabolism.

Multiple drug resistance, and hence tumor cell selection, outgrowth, and relapse, leads to morbidity and mortality in cancer patients. Candidate libraries can be introduced into tumor cell lines (primary and cultured) that have demonstrated specific or multiple drug resistance. Bioactive peptides can then be identified which confer drug sensitivity when the cells are exposed to the drug of interest, or to drugs used in combination chemotherapy. The readout can be the onset of apoptosis in these cells, membrane permeability changes, the release of intracellular ions and fluorescent markers. The cells in which multidrug resistance involves membrane transporters can be preloaded with fluorescent transporter substrates, and selection carried out for peptides which block the normal efflux of fluorescent drug from these cells. Candidate libraries are particularly suited to screening for peptides which reverse poorly characterized or recently discovered intracellular mechanisms of resistance or mechanisms for which few or no chemosensitizers currently exist, such as mechanisms involving LRP (lung resistance protein). This protein has been implicated in multidrug resistance in ovarian carcinoma, metastatic malignant melanoma, and acute myeloid leukemia. Particularly interesting examples include screening for agents which reverse more than one important resistance mechanism in a single cell, which occurs in a subset of the most drug resistant cells, which are also important targets. Applications would include screening for peptide inhibitors of both MRP (multidrug resistance related protein) and LRP for treatment of resistant cells in metastatic melanoma, for inhibitors of both p-glycoprotein and LRP in acute myeloid leukemia, and for inhibition (by any mechanism) of all three proteins for treating pan-resistant cells.

In a preferred embodiment, the present methods are useful in improving the performance of existing or developmental drugs. First pass metabolism of orally administered drugs limits their oral bioavailability, and can result in diminished efficacy as well as the need to administer more drug for a desired effect. Reversible inhibitors of enzymes involved in first pass metabolism may thus be a useful adjunct enhancing the efficacy of these drugs. First pass metabolism occurs in the liver, thus inhibitors of the corresponding catabolic enzymes may enhance the effect of the cognate drugs. Reversible inhibitors would be delivered at the same time as, or slightly before, the drug of interest. Screening of candidate libraries in hepatocytes for inhibitors (by any mechanism, such as protein downregulation as well as a direct inhibition of activity) of particularly problematical isozymes would be of interest. These include the CYP3A4 isozymes of cytochrome P450, which are involved in the first pass metabolism of the anti-HIV drugs saquinavir and indinavir. Other applications could include reversible inhibitors of UDP-glucuronyltransferases, sulfotransferases, N-acetyltransferases, epoxide hydrolases, and glutathione S-transferases, depending on the drug. Screens would be done in cultured hepatocytes or liver microsomes, and could involve antibodies recognizing the specific modification performed in the liver, or co-cultured readout cells, if the metabolite had a different bioactivity than the untransformed drug. The enzymes modifying the drug would not necessarily have to be known, if screening was for lack of alteration of the drug.

In a preferred embodiment, the present methods are useful in immunobiology, inflammation, and allergic response applications. Selective regulation of T lymphocyte responses is a desired goal in order to modulate immune-mediated diseases in a specific manner. Candidate libraries can be introduced into specific T cell subsets (TH1, TH2, CD4+, CD8+, and others) and the responses which characterize those subsets (cytokine generation, cytotoxicity, proliferation in response to antigen being presented by a mononuclear leukocyte, and others) modified by members of the library. Agents can be selected which increase or diminish the known T cell subset physiologic response. This approach will be useful in any number of conditions, including: 1) autoimmune diseases where one wants to induce a tolerant state (select a peptide that inhibits T cell subset from recognizing a self-antigen bearing cell); 2) allergic diseases where one wants to decrease the stimulation of IgE producing cells (select peptide which blocks release from T cell subsets of specific B-cell stimulating cytokines which induce switch to IgE production); 3) in transplant patients where one wants to induce selective immunosuppression (select peptide that diminishes proliferative responses of host T cells to foreign antigens); 4) in lymphoproliferative states where one wants to inhibit the growth or sensitize a specific T cell tumor to chemotherapy and/or radiation; 5) in tumor surveillance where one wants to inhibit the killing of cytotoxic T cells by Fas ligand bearing tumor cells; and 5) in T cell mediated inflammatory diseases such as Rheumatoid arthritis, Connective tissue diseases (SLE), Multiple sclerosis, and inflammatory bowel disease, where one wants to inhibit the proliferation of disease-causing T cells (promote their selective apoptosis) and the resulting selective destruction of target tissues (cartilage, connective tissue, oligodendrocytes, gut endothelial cells, respectively).

Regulation of B cell responses will permit a more selective modulation of the type and amount of immunoglobulin made and secreted by specific B cell subsets. Candidate libraries can be inserted into B cells and bioactive peptides selected which inhibit the release and synthesis of a specific immunoglobulin. This may be useful in autoimmune diseases characterized by the overproduction of auto antibodies and the production of allergy causing antibodies, such as IgE. Agents can also be identified which inhibit or enhance the binding of a specific immunoglobulin subclass to a specific antigen either foreign of self. Finally, agents can be selected which inhibit the binding of a specific immunoglobulin subclass to its receptor on specific cell types.

Similarly, agents which affect cytokine production may be selected, generally using two cell systems. For example, cytokine production from macrophages, monocytes, etc. may be evaluated. Similarly, agents which mimic cytokines, for example erythropoetin and IL1-17, may be selected, or agents that bind cytokines such as TNF-α, before they bind their receptor.

Antigen processing by mononuclear leukocytes (ML) is an important early step in the immune system's ability to recognize and eliminate foreign proteins. Candidate agents can be inserted into ML

cell lines and agents selected which alter the intracellular processing of foreign peptides and sequence of the foreign peptide that is presented to T cells by MLs on their cell surface in the context of Class II MHC. One can look for members of the library that enhance immune responses of a particular T cell subset (for example, the peptide would in fact work as a vaccine), or look for a library member that binds more tightly to MHC, thus displacing naturally occurring peptides, but nonetheless the agent would be less immunogenic (less stimulatory to a specific T cell clone). This agent would in fact induce immune tolerance and/or diminish immune responses to foreign proteins. This approach could be used in transplantation, autoimmune diseases, and allergic diseases.

The release of inflammatory mediators (cytokines, leukotrienes, prostaglandins, platelet activating factor, histamine, neuropeptides, and other peptide and lipid mediators) is a key element in maintaining and amplifying aberrant immune responses. Candidate libraries can be inserted into MLs, mast cells, eosinophils, and other cells participating in a specific inflammatory response, and bioactive peptides selected which inhibit the synthesis, release and binding to the cognate receptor of each of these types of mediators.

In a preferred embodiment, the present methods are useful in biotechnology applications. Candidate library expression in mammalian cells can also be considered for other pharmaceutical-related applications, such as modification of protein expression, protein folding, or protein secretion. One such example would be in commercial production of protein pharmaceuticals in CHO or other cells. Candidate libraries resulting in bioactive peptides which select for an increased cell growth rate (perhaps peptides mimicking growth factors or acting as agonists of growth factor signal transduction pathways), for pathogen resistance (see previous section), for lack of sialylation or glycosylation (by blocking glycotransferases or rerouting trafficking of the protein in the cell), for allowing growth on autoclaved media, or for growth in serum free media, would all increase productivity and decrease costs in the production of protein pharmaceuticals.

Random peptides displayed on the surface of circulating cells can be used as tools to identify organ, tissue, and cell specific peptide targeting sequences. Any cell introduced into the bloodstream of an animal expressing a library targeted to the cell surface can be selected for specific organ and tissue targeting. The bioactive peptide sequence identified can then be coupled to an antibody, enzyme, drug, imaging agent or substance for which organ targeting is desired.

Other agents which may be selected using the present invention include: 1) agents which block the activity of transcription factors, using cell lines with reporter genes; 2) agents which block the interaction of two known proteins in cells, using the absence of normal cellular functions, the mammalian two hybrid system or fluorescence resonance energy transfer mechanisms for detection;

and 3) agents may be identified by tethering a random peptide to a protein binding region to allow interactions with molecules sterically close, i.e. within a signalling pathway, to localize the effects to a functional area of interest.

In a preferred embodiment, the bioactive peptide may also be used in gene therapy. In gene therapy applications, genes encoding the peptide are introduced into cells in order to achieve *in vivo* synthesis of a therapeutically effective genetic product. "Gene therapy" includes both conventional gene therapy where a lasting effect is achieved by a single treatment, and the administration of gene therapeutic agents, which involves the one time or repeated administration of a therapeutically effective DNA or mRNA.

There are a variety of techniques available for introducing nucleic acids into viable cells. The techniques vary depending upon whether the nucleic acid is transferred into cultured cells *in vitro*, or *in vivo* in the cells of the intended host. Techniques suitable for the transfer of nucleic acid into mammalian cells *in vitro* include the use of liposomes, electroporation, microinjection, cell fusion, DEAE-dextran, the calcium phosphate precipitation method, etc. The currently preferred *in vivo* gene transfer techniques include transfection with viral (typically retroviral) vectors and viral coat protein-liposome mediated transfection [Dzau et al., Trends in Biotechnology 11:205-210 (1993)]. In some situations it is desirable to provide the nucleic acid source with an agent that targets the target cells, such as an antibody specific for a cell surface membrane protein or the target cell, a ligand for a receptor on the target cell, etc. Where liposomes are employed, proteins which bind to a cell surface membrane protein associated with endocytosis may be used for targeting and/or to facilitate uptake, e.g. capsid proteins or fragments thereof tropic for a particular cell type, antibodies for proteins which undergo internalization in cycling, proteins that target intracellular localization and enhance intracellular half-life. The technique of receptor-mediated endocytosis is described, for example, by Wu et al., J. Biol. Chem. 262:4429-4432 (1987); and Wagner et al., Proc. Natl. Acad. Sci. U.S.A. 87:3410-3414 (1990). For review of gene marking and gene therapy protocols see Anderson et al., Science 256:808-813 (1992).

Alternatively, an *ex vivo* approach can be used in which a cell excreting a therapeutically effective peptide may be transplanted into an individual, for the constant or regulated systemic delivery of the peptide.

The pharmaceutical compositions of the present invention comprise a compound in a form suitable for administration to a patient. In the preferred embodiment, the pharmaceutical compositions are in a water soluble form, such as being present as pharmaceutically acceptable salts, which is meant to

include both acid and base addition salts. "Pharmaceutically acceptable acid addition salt" refers to those salts that retain the biological effectiveness of the free bases and that are not biologically or otherwise undesirable, formed with inorganic acids such as hydrochloric acid, hydrobromic acid, sulfuric acid, nitric acid, phosphoric acid and the like, and organic acids such as acetic acid, propionic acid, glycolic acid, pyruvic acid, oxalic acid, maleic acid, malonic acid, succinic acid, fumaric acid, tartaric acid, citric acid, benzoic acid, cinnamic acid, mandelic acid, methanesulfonic acid, ethanesulfonic acid, p-toluenesulfonic acid, salicylic acid and the like. "Pharmaceutically acceptable base addition salts" include those derived from inorganic bases such as sodium, potassium, lithium, ammonium, calcium, magnesium, iron, zinc, copper, manganese, aluminum salts and the like. Particularly preferred are the ammonium, potassium, sodium, calcium, and magnesium salts. Salts derived from pharmaceutically acceptable organic non-toxic bases include salts of primary, secondary, and tertiary amines, substituted amines including naturally occurring substituted amines, cyclic amines and basic ion exchange resins, such as isopropylamine, trimethylamine, diethylamine, triethylamine, tripropylamine, and ethanolamine.

The compounds can be formulated using pharmaceutically acceptable carriers into dosages suitable for oral administration. Such carriers enable the compounds of the invention to be formulated as tablets, pills, capsules, liquids, gels, syrups, slurries, and the like for oral ingestion.

The administration of the bioactive peptides of the present invention, preferably in the form of a sterile aqueous solution, can be done in a variety of ways, including, but not limited to, orally, subcutaneously, intravenously, intranasally, transdermally, intraperitoneally, intramuscularly, intrapulmonary, vaginally, rectally, or intraocularly. In some instances, for example, in the treatment of wounds, inflammation, etc., the peptide may be directly applied as a solution or spray. Depending upon the manner of introduction, the pharmaceutical composition may be formulated in a variety of ways. The concentration of the therapeutically active peptide in the formulation may vary from about 0.1 to 100 weight %.

The pharmaceutical compositions may also include one or more of the following: carrier proteins such as serum albumin; buffers; fillers such as microcrystalline cellulose, lactose, corn and other starches; binding agents; sweeteners and other flavoring agents; coloring agents; and polyethylene glycol. Additives are well known in the art, and are used in a variety of formulations.

The following examples serve to more fully describe the manner of using the above-described invention, as well as to set forth the best modes contemplated for carrying out various aspects of the invention. It is understood that these examples in no way serve to limit the true scope of this invention,

but rather are presented for illustrative purposes. All references cited herein are incorporated by reference.

EXAMPLES

Example 1

Isolation of Inteins with Altered Cyclization Activity

A fluorescent reporter system was designed for quantifying intein cyclization. GFP was split at the loop 3 junction and the translational order of the N and C-terminal fragments were reversed (Figure 12A). The termini were held together by a glycine-serine linker. In some constructs, one-half of the myc epitope was fused onto either side of theloop 3 junction (Figure 12A). The resulting GFP molecules were positioned with an intein scaffold comprising either wild-type or a mutant intein (Figure 12C).

Mutant intein sequences obtained using PCR mutagenesis were screened for activity by FACS sorting for increases in fluorescence.. Western blot analysis of several other mutants is shown in Figure 13. In Figure 13, several of the mutants had cyclization efficiencies greater than theparental staring intein, J3.

Example 2

Biasing a Cyclic Peptide to Reduce the Number of Conformers

To test the effects of a fixed proline in a cyclic 7mer, the conformation space of the 7mer cyclic peptide RGDGWS, containing two flexible glycines was compared with that of cyclic RGPGWS using quenced molecular dynamics calculations (O'Connor, et al., (1992) J. Med. Chem., 35:2870-81); Mackay, et al., (1989) "The role of energy minimization in simulation stragegies of biomolecular systems", In *Prediction of Protein Structure and the Principles of Protein Conformation*, Fasman, G., ed., New York, Plemum Press, pp. 317-358).

The lowest 5 kcal energy conformers were collected from a total of at least 10,000 individual conformers obtained from multiple molecular dynamics trajectories, and compared with each other using the backbone amino acids by overlaying the structures and calculating the root mean square deviation of these atoms in the best fit overlay using InsightII (Molecular Simulations Inc.).

An example of the cluster graph of the lowest energy conformers for each peptide is shown in Figures 15 and 16. The root mean square deviation (RMSD, Å) is coded by color, with very similar conformers (RMSD ≤ Å) in yellow, still highly similar conformers (RMSD between 1-2 Å)in white, similar conformers (RMSD between 2-3 Å) in blue, less similar conformers (RMSD between 3-4 Å) in red, and dissimilar conformers in black (not shown).

For the cyclic peptide SRGDGWS, shown in Figure 15 (srgdgwsLowest5A.ps), there were 62 low energy conformers. There was one family of very similar conformers (yellow square at bottom left) and two families of quite similar conformers in yellow/white, one roughly in the middle of the graph, and one (with only moderately similar conformers) near the top right corner. These comprised approximately 20 of the 62 conformers. The rest of the low energy conformers were not very similar to each other, and much of the graph is red or black. Backbone overlaid conformers from most similar family, No. 1, are shown at the lower left. In the lower middle, is family No. 2. these conformers, when overlaid are clearly not similar. Conformers in family No. 3 (lower right), are rather heterogeneous, although not as much as those from the red and black regions of the graph.

For the cyclic peptide SRGPGWS, representing the substitution of pro for asp 4, the graph of the lowest energy conformers looks quite different (Figure 16; srgpgwsLowest5B.ps). There is a much larger family of very similar conformers (lower left of graph, family No. 1, conformers 1-26). Family No. 2 also has very similar conformers, although they are all different from family No. 1. Even family No. 3, representing over two thirds of all low energy conformers (frames 1-59) contains conformers that are similar enough to give a blurred donut appearance. Thus, substitution of a single pro for another residue (asp in this case) clearly freezes out two additional families of conformers. As this peptide has two glycines, the effect of proline on conformational narrowing of cyclic peptides with 1 or 0 glycines may be more profound.